AMERICAN MATHEMATICAL SOCIETY
COLLOQUIUM PUBLICATIONS, VOLUME XI

THE THEORY OF APPROXIMATION

BY

DUNHAM JACKSON

PROFESSOR OF MATHEMATICS IN THE
UNIVERSITY OF MINNESOTA

NEW YORK
PUBLISHED BY THE
AMERICAN MATHEMATICAL SOCIETY
501 WEST 116TH STREET
1930

Photographically
reproduced *1951*
EDWARDS BROTHERS, INC.
Lithoprinters
ANN ARBOR, MICHIGAN

PREFACE

The title of this volume is an abbreviation for the more properly descriptive one: "Topics in the theory of approximation". It is a brief essay in a field on which an encyclopedia might be written. On the personal side, it is an account of certain aspects and ramifications of a problem to which I was introduced at an early stage, and which has given direction to my reading and study ever since.

One day about twenty years ago I was admitted to the study of Professor Landau, seeking advice as to a subject for a thesis. After some preliminary inquiries as to my experience and preferences, he handed me a long sheet of paper, and directed me to take notes as he enumerated some dozen or fifteen topics in various fields of analysis and number theory, with a few words of explanation of each. He told me to think about them for a few days, and to select one of them, or any other problem of my own choosing, with the single reservation that I should *not* prove Fermat's theorem, an injunction which I have observed faithfully. Guided partly by natural inclination, perhaps, and partly by recollection of a course on methods of approximation which I had taken with Professor Bôcher a few years earlier, I committed myself to one of the topics which Landau had proposed, an investigation of the degree of approximation with which a given continuous function can be represented by a polynomial of given degree. When I reported my choice, he said meditatively, in words which I remember vividly in substance, if not perfectly as to idiom: "Das ist ein schönes Thema, ich beneide Sie um das Thema ... Nein, ich beneide Sie nicht, aber es ist ein wunderschönes Thema!" It is in fact a problem which admits a surprising variety of interesting developments on

its own account, and offers a natural avenue of approach to a number of fields of still broader importance.

Although delayed in its completion by the conflict of other duties, the following exposition is substantially in the form in which it was projected at the time of the Colloquium lectures in 1925, and presented in abstract in the lectures themselves. One section, on the vector analysis of function space, originally designed for inclusion in the Colloquium, has meanwhile been published separately instead. The sections which had been written at full length in September, 1925 — practically the whole of the first chapter, parts of the second, and most of the third — have been left unchanged, except in minor details. The elementary account of Legendre series in Chapter I, for example, was written before the appearance of the admirable article on the subject by M. H. Stone in vol. 27 of the Annals of Mathematics. A few other articles published since 1925 are mentioned in the text.

For the most part, however, citations of the literature have been omitted. The preparation of a really adequate bibliography would have been a task of such magnitude as to delay the publication indefinitely. References to some of the most important papers of not too recent date are contained in my thesis (Göttingen, 1911) and in my report on *The general theory of approximation by polynomials and trigonometric sums* in vol. 27 of the Bulletin of the American Mathematical Society. Among publications in book form supplementing the material given here, mention should be made of Borel's *Leçons sur les fonctions de variables réelles et les développements en séries de polynomes,* de la Vallée Poussin's *Leçons sur l'approximation des fonctions d'une variable réelle,* and S. Bernstein's *Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle,* all appearing in the Borel series. As to the content of these lectures themselves, there are many points where it would be difficult now to recall the original sources either of specific results and proofs or of suggestions as to method. To the extent that the work is my own, some

parts have been published previously, in my thesis, in various articles in the Transactions of the American Mathematical Society, and elsewhere; other parts are now offered in print for the first time. Numerous detailed acknowledgments, not repeated here, have been made in the pages of the earlier publications. In connection with Chapter IV, reference should still be made to the work of Faber on trigonometric interpolation in his memoir *Über stetige Funktionen (zweite Abhandlung)* in vol. 69 of the Mathematische Annalen. My acquaintance with the statistical formulas discussed in Chapter V, which might have come from any of a variety of sources, was in fact mostly obtained from Yule's *Introduction to the Theory of Statistics.* The lemma on which the method of Chapter III depends is derived from the most important single memoir in the literature on degree of approximation, S. Bernstein's epoch-making prize essay of 1912, with which the present work also has other points of contact. And in conclusion it should be said that my study of the problem has been dominated from the beginning not only by the influence of my own teachers, but also by the writings of Lebesgue and de la Vallée Poussin.

October 1, 1929

DUNHAM JACKSON

CONTENTS

CHAPTER I

CONTINUOUS FUNCTIONS

CHAPTER II

DISCONTINUOUS FUNCTIONS: FUNCTIONS OF LIMITED VARIATION; ARITHMETIC MEANS

CHAPTER III

THE PRINCIPLE OF LEAST SQUARES AND ITS GENERALIZATIONS

CHAPTER IV

TRIGONOMETRIC INTERPOLATION

CHAPTER V

INTRODUCTION TO THE GEOMETRY OF FUNCTION SPACE

CHAPTER I

CONTINUOUS FUNCTIONS

Introduction

Weierstrass first enunciated the theorem that an arbitrary continuous function can be approximately represented by a polynomial with any assigned degree of accuracy. The theorem may be stated with precision in the following form:

If $f(x)$ is a given function, continuous for $a \leq x \leq b$, and if ε is a given positive quantity, it is always possible to define a polynomial $P(x)$ such that

$$|f(x) - P(x)| < \varepsilon$$

for $a \leq x \leq b$.

To Weierstrass is due also the corresponding theorem on approximation by means of trigonometric sums:

If $f(x)$ is a given function of period 2π, continuous for all real values of x, and if ε is a given positive quantity, it is always possible to define a trigonometric sum $T(x)$ such that

$$|f(x) - T(x)| < \varepsilon$$

for all real values of x.

By a *polynomial* is meant an expression of the form

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n.$$

This expression will be said to represent a polynomial *of the nth degree,* not only when a_n is different from zero, but, in distinction from the usage which prevails in some parts of algebra, also when $a_n = 0$. That is to say, the words "polynomial of the nth degree" will be used in place of the longer expression "polynomial of the nth degree at most". Even the case of identical vanishing is not excluded. A *trig-*

onometric sum, or more specifically a trigonometric sum *of the nth order,* is an expression of the form

$$a_0 + a_1 \cos x + a_2 \cos 2x + \cdots + a_n \cos nx$$
$$+ b_1 \sin x + b_2 \sin 2x + \cdots + b_n \sin nx.$$

The definition is inclusive once more; the simultaneous vanishing of a_n and b_n is not ruled out.

These two types of approximating function show a persistent and fundamental similarity in their behavior, on which differences of more or less significance are from time to time superimposed. Simplicity of statement and proof will favor sometimes one and sometimes the other.

It is readily seen that the number of terms required to yield a specified degree of approximation, or, under the converse aspect, the degree of approximation attainable with a specified number of terms, will be related to the properties of continuity of $f(x)$. It is the purpose of the next pages to trace out this relationship in some detail.

1. Approximation by trigonometric sums

For a considerable body of results, the following theorem may be regarded as fundamental:

THEOREM I. *If $f(x)$ is a function of period 2π, such that*

$$|f(x_2) - f(x_1)| \leq \lambda |x_2 - x_1|$$

for all real values of x_1 and x_2, λ being a constant, there will exist for every positive integral value of n a trigonometric sum $T_n(x)$, of the nth order, such that, for all real values of x,

$$|f(x) - T_n(x)| \leq \frac{K\lambda}{n},$$

where K is an absolute constant, depending neither on x, nor on n, nor on λ, nor on any further specification with regard to the function $f(x)$.

In the proof of the theorem, use will be made of the following

LEMMA. *If m is a positive integer, the expression*

$$\frac{\sin^4(mx/2)}{\sin^4(x/2)}$$

is a trigonometric sum in x, of order $2m-2$.

Because of the identity

$$\cos px \cos qx = \frac{1}{2}[\cos(p+q)x + \cos(p-q)x]$$

and the others of similar type, it is seen at once that the product of two trigonometric sums, of orders n_1 and n_2 respectively, is a trigonometric sum of order n_1+n_2. It is sufficient for the purpose in hand, therefore, to recall any one of the numerous proofs of the well-known fact that

$$\frac{\sin^2(mx/2)}{\sin^2(x/2)} = \frac{1-\cos mx}{1-\cos x}$$

is a trigonometric sum of order $m-1$; its square will then be a sum of order $2m-2$. The fact that $1-\cos mx$ is equal to the product of $1-\cos x$ by a trigonometric sum of order $m-1$ appears, for example, from the identities

$$1-\cos mx = \sum_{p=0}^{m-1}[\cos px - \cos(p+1)x],$$

$$\begin{aligned}\cos px - \cos(p+1)x \\ &= (1-\cos x) - \sum_{q=1}^{p}[\cos(q-1)x - 2\cos qx \\ &\qquad\qquad + \cos(q+1)x],\end{aligned}$$

$$\begin{aligned}\cos(q-1)x - 2\cos qx + \cos(q+1)x \\ &= [\cos(q-1)x + \cos(q+1)x] - 2\cos qx \\ &= 2\cos qx\cos x - 2\cos qx \\ &= -2\cos qx(1-\cos x).\end{aligned}$$

To proceed with the proof of the theorem, let

$$F_m(u) = \left[\frac{\sin mu}{m\sin u}\right]^4, \quad I_m(x) = h_m\int_{-\pi/2}^{\pi/2} f(x+2u)\,F_m(u)\,du,$$

where m is any positive integer, and h_m is defined by the equation

$$\frac{1}{h_m} = \int_{-\pi/2}^{\pi/2} F_m(u)\,du.$$

By means of the substitution $x+2u=v$, the expression for $I_m(x)$ is transformed into

$$\frac{1}{2}\,h_m \int_{x-\pi}^{x+\pi} f(v)\,F_m\left[\frac{1}{2}(v-x)\right]dv.$$

Both factors in the last integrand have the period 2π with regard to v, so that the value of the integral is unchanged if the interval of integration is replaced by any other interval of length 2π. In particular,

$$I_m(x) = \frac{1}{2}\,h_m \int_{-\pi}^{\pi} f(v)\,F_m\left[\frac{1}{2}(v-x)\right]dv.$$

The expression $F_m[\frac{1}{2}(v-x)]$, by the Lemma above, is a trigonometric sum of order $2m-2$ in $(v-x)$, and may be regarded as a trigonometric sum of the same order in x, with coefficients which are trigonometric functions of v. The whole integrand is a trigonometric sum in x with coefficients which are continuous functions of v, and $I_m(x)$ therefore is a trigonometric sum of order $2m-2$ in x, with constant coefficients. The proof that this sum is an approximate representation of $f(x)$, when m is large, will be based on the original representation of $I_m(x)$.

Let the equation defining h_m be multiplied by $h_m f(x)$. Since $f(x)$ is a constant as far as u is concerned, it may be placed under the sign of integration, so that

$$f(x) = h_m \int_{-\pi/2}^{\pi/2} f(x)\,F_m(u)\,du.$$

Consequently

$$I_m(x) - f(x) = h_m \int_{-\pi/2}^{\pi/2} [f(x+2u)-f(x)]\,F_m(u)\,du.$$

By the hypothesis imposed on $f(x)$,

$$|f(x+2u)-f(x)| \leqq 2\lambda\,|u|.$$

Hence

$$|I_m(x)-f(x)| \leqq 2\lambda h_m \int_{-\pi/2}^{\pi/2} |u|\, F_m(u)\, du,$$

or, since $F_m(u)$ and $|u|\, F_m(u)$ are even functions of u,

$$|I_m(x)-f(x)| \leqq 4\lambda h_m \int_0^{\pi/2} u\, F_m(u)\, du = 2\lambda \frac{\int_0^{\pi/2} u\, F_m(u)\, du}{\int_0^{\pi/2} F_m(u)\, du}.$$

To anticipate the conclusion of the proof, let

$$c_1 = \int_0^{\pi/2} \frac{\sin^4 t}{t^4}\, dt, \qquad c_2 = \int_0^{\infty} \frac{\sin^4 t}{t^3}\, dt.$$

These quantities are merely numerical constants. Is is clear that each integrand approaches a limit for $t = 0$, and that the improper integral defining c_2 is convergent.

By the use of the fact that $0 < \sin u < u$ for $0 < u \leqq \pi/2$, and the substitution $mu = t$, it is recognized that

$$\int_0^{\pi/2} F_m(u)\, du > \int_0^{\pi/2} \left[\frac{\sin mu}{mu}\right]^4 du = \frac{1}{m} \int_0^{m\pi/2} \frac{\sin^4 t}{t^4}\, dt$$
$$\geqq \frac{1}{m} \int_0^{\pi/2} \frac{\sin^4 t}{t^4}\, dt = \frac{c_1}{m}.$$

On the other hand, $(\sin u)/u$ decreases monotonically as u goes from 0 to $\pi/2$, so that

$$\frac{\sin u}{u} > \frac{\sin(\pi/2)}{(\pi/2)} = \frac{2}{\pi}, \qquad \frac{1}{\sin u} < \frac{\pi}{2} \cdot \frac{1}{u}$$

throughout the interior of this interval. Hence

$$\int_0^{\pi/2} u\, F_m(u)\, du \leqq \left(\frac{\pi}{2}\right)^4 \int_0^{\pi/2} u \left[\frac{\sin mu}{mu}\right]^4 du$$
$$= \frac{1}{m^2}\left(\frac{\pi}{2}\right)^4 \int_0^{m\pi/2} \frac{\sin^4 t}{t^3}\, dt$$
$$< \frac{1}{m^2}\left(\frac{\pi}{2}\right)^4 \int_0^{\infty} \frac{\sin^4 t}{t^3}\, dt = \left(\frac{\pi}{2}\right)^4 \frac{c_2}{m^2}.$$

From these relations it follows that

$$|I_m(x)-f(x)| \leqq \frac{\pi^4 c_2 \lambda}{8 c_1 m}.$$

Now let n be an arbitrary integer, and let m be taken equal to $\frac{1}{2}n+1$ or $\frac{1}{2}(n+1)$, according as n is even or odd. In either case, $2m-2 \leqq n < 2m$. Let the corresponding expression $I_m(x)$ be denoted by $T_n(x)$. Then $T_n(x)$ is a trigonometric sum of the nth order (it will be remembered that this is understood to mean of the nth order *at most*, according to the more usual terminology), and, since $1/m < 2/n$,

$$|I_m(x) - f(x)| \leq \frac{\pi^4 c_2 \lambda}{4 c_1 n} = \frac{K\lambda}{n}.$$

if K is taken equal to $\pi^4 c_2/(4c_1)$. Thus the proof of the theorem is completed.

So much has been conceded to simplicity of outline, in building up the above inequalities, that the final upper limits are quite unnecessarily large, giving little indication of the actual magnitude of the quantities that precede. It will add a little to the definiteness of the conclusion to point out that $c_1 > (2/\pi)^3$, since $(\sin t)/t > 2/\pi$ throughout the interior of the interval of integration, while

$$c_2 = \int_0^1 t\cdot\frac{\sin^4 t}{t^4}\,dt + \int_1^\infty \frac{\sin^4 t}{t^3}\,dt < \int_0^1 t\,dt + \int_1^\infty \frac{dt}{t^3} = 1,$$

so that

$$K = \frac{\pi^4 c_2}{4c_1} < \frac{\pi^7}{32} < 100.$$

With more attention to detail, however, the estimate can be cut very much closer. The theorem is actually true with $K=3$, instead of the value adopted above, or even with a somewhat smaller value of K. On the other hand, it can be shown that the statement is *not* generally true with a value of K smaller than $\pi/2$.

To pass on to a more general theorem, let $f(x)$ be an arbitrary continuous function of period 2π, and let $\omega(\delta)$ be the maximum of $|f(x_2)-f(x_1)|$ for $|x_2-x_1| \leq \delta$. The function $\omega(\delta)$ has been called by de la Vallée Poussin the *modulus of continuity* of $f(x)$. With the word *maximum* replaced by *least upper bound*, it can be defined for any

bounded function, whether continuous or not. The characteristic property of a uniformly continuous function is that $\lim_{\delta=0} \omega(\delta) = 0$.

Let $\varphi(x)$ be the continuous function of period 2π which takes on the same values as $f(x)$ at the points

$$-\pi, \quad -\pi+\frac{2\pi}{n}, \quad -\pi+\frac{4\pi}{n}, \quad \ldots, \quad \pi-\frac{2\pi}{n}, \quad \pi,$$

and is linear from each point of this set to the next. The graph of $\varphi(x)$ is a broken line, no segment of which has a slope greater than $\omega(2\pi/n)/(2\pi/n)$ in absolute value. In analytical language, $\varphi(x)$ satisfies the hypothesis of Theorem I, with

$$\lambda = \frac{\omega(2\pi/n)}{2\pi/n}.$$

For every positive integral value of n, therefore, there is a trigonometric sum $T_n(x)$, of the nth order, such that

$$|\varphi(x)-T_n(x)| \leqq \frac{K}{2\pi}\,\omega\left(\frac{2\pi}{n}\right).$$

On the other hand, any specified value of x differs by less than $2\pi/n$ from one of those for which f and φ are by definition equal to each other; neither $f(x)$ nor $\varphi(x)$ can differ by more than $\omega(2\pi/n)$ from the corresponding common value; and hence

$$|f(x)-\varphi(x)| \leqq 2\,\omega\left(\frac{2\pi}{n}\right)$$

for all values of x. If the quantity $K/(2\pi)+2$ is denoted by K', the last two inequalities may be combined to yield the following statement:

THEOREM II. *If $f(x)$ is a continuous function of period 2π, with modulus of continuity $\omega(\delta)$, there exists for every positive integral value of n a trigonometric sum $T_n(x)$, of the nth order, such that, for all real values of x,*

$$|f(x)-T_n(x)| < K'\,\omega\left(\frac{2\pi}{n}\right),$$

where K' is an absolute constant.

While this theorem is applicable to any continuous function, it involves the modulus of continuity in the inequality which forms the essence of its conclusion. It can be shown that the assignment of an outer limit of error for an arbitrary continuous function, without some dependence on properties of the function beyond the mere fact of its continuity, is impossible.

Since $\lim_{n=\infty} \omega(2\pi/n) = 0$, it is to be noted that Theorem II includes one of the theorems of Weierstrass to which reference was made in the opening lines of the chapter.

In preparation for the next developments, there is occasion to examine more closely the proof that was given above for Theorem I. It will be recalled that to an arbitrary positive integer n a second positive integer m was assigned, in terms of which a function $F_m(u)$ was constructed; and a trigonometric sum $T_n(x)$, yielding an approximate representation of the given function $f(x)$, was defined as equal to an expression which could be reduced to the form

$$\frac{1}{2} h_m \int_{-\pi}^{\pi} f(v)\, F_m\left[\frac{1}{2}(v-x)\right] dv,$$

h_m being independent of x. A lemma stated essentially that $F_m(\frac{1}{2}u)$ is a trigonometric sum in u, of order $2m-2 \leqq n$. It is possible therefore to write $F_m[\frac{1}{2}(v-x)]$ in the form

$$\frac{1}{2} A_{n0} + \sum_{k=1}^{n} [A_{nk} \cos k(v-x) + B_{nk} \sin k(v-x)].$$

When the above expression for $T_n(x)$ is expanded as a trigonometric sum in x, the constant term is

$$\frac{1}{4} h_m A_{n0} \int_{-\pi}^{\pi} f(v)\, dv,$$

and is zero if the last integral vanishes, an observation which will presently be important, for the reason that the indefinite integral of a trigonometric sum without constant term is itself a trigonometric sum, while this is not the case if the sum to be integrated has a constant term different from zero.

It may be pointed out in this connection, though it is not essential to the main argument, that the coefficients B_{nk} are all zero. This can be inferred from an elementary theorem on trigonometric sums, since $F_m(\frac{1}{2}u)$ is an even function of u, and is also directly apparent on inspection of the proof of the lemma. If a_k, b_k are the Fourier coefficients of $f(x)$:

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(v)\cos kv\,dv, \quad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(v)\sin kv\,dv,$$

and if $\frac{1}{2}\pi h_m A_{nk}$ is denoted by d_{nk}, it is seen that

$$T_n(x) = \frac{1}{2} d_{n0}\, a_0 + \sum_{k=1}^{n} d_{nk}(a_k \cos kx + b_k \sin kx).$$

As the d's are independent of the function to be represented, the calculation of the successive expressions $T_n(x)$ amounts to a method of summation of the Fourier series for $f(x)$.

To return from the digression of the last paragraph, let $f(x)$ be a function of period 2π, which has everywhere a continuous derivative $f'(x)$. For a particular value of n, let $t_n'(x)$ be a trigonometric sum of the nth order, without constant term:

$$t_n'(x) = \sum_{k=1}^{n} (\alpha_k \cos kx + \beta_k \sin kx),$$

and let ϵ_n be a constant such that

$$|f'(x) - t_n'(x)| \leq \epsilon_n$$

for all values of x. Let $t_n(x)$ be the trigonometric sum, without constant term, which has $t_n'(x)$ for its derivative:

$$t_n(x) = \sum_{k=1}^{n} \left(\frac{\alpha_k}{k} \sin kx - \frac{\beta_k}{k} \cos kx\right),$$

and let $r_n(x) = f(x) - t_n(x)$. *Then $r_n(x)$ has the period 2π, and, since $|r_n'(x)| \leq \epsilon_n$, satisfies the conditions imposed on $f(x)$ in the hypothesis of Theorem* I, *with $\lambda = \epsilon_n$.* Hence there

exists a trigonometric sum of the nth order, which may be denoted by $\tau_n(x)$, such that

$$|r_n(x) - \tau_n(x)| \leqq \frac{K\varepsilon_n}{n}.$$

If $T_n(x) = t_n(x) + \tau_n(x)$, then $f(x) - T_n(x) = r_n(x) - \tau_n(x)$, and

$$|f(x) - T_n(x)| \leqq \frac{K\varepsilon_n}{n}.$$

From the existence of an approximation for $f'(x)$ it has been possible to draw an important inference with regard to the approximation of $f(x)$. If $f(x)$ is itself the derivative of a function of period 2π, so that the integral of $f(x)$ over an interval of length 2π is zero, it follows that

$$\int_{-\pi}^{\pi} r_n(x)\,dx = \int_{-\pi}^{\pi} f(x)\,dx - \int_{-\pi}^{\pi} t_n(x)\,dx = 0,$$

whence, according to the second paragraph preceding, the sum $\tau_n(x)$ given by the proof of Theorem I as an approximation for $r_n(x)$ will have no constant term. So the constant term in the present $T_n(x)$, defined in terms of this $\tau_n(x)$, will be zero likewise.

The way has now been prepared for a demonstration of

THEOREM III. *If $f(x)$ is a function of period 2π, having a pth derivative $f^{(p)}(x)$ such that*

$$|f^{(p)}(x_2) - f^{(p)}(x_1)| \leqq \lambda |x_2 - x_1|$$

for all real values of x_1 and x_2, λ being a constant, there will exist for every positive integral value of n a trigonometric sum $T_n(x)$, of the nth order, such that, for all real values of x,

$$|f(x) - T_n(x)| \leqq \frac{K^{p+1}\lambda}{n^{p+1}},$$

where K is the absolute constant found in the proof of Theorem I.

It is to be noticed that the argument is based on the explicit construction of the approximating sum in Theorem I,

and it is not clear that a smaller value of K in that theorem, justified by some different method, would necessarily be applicable here. The value $K = 3$, previously mentioned in connection with Theorem I, results from the same explicit construction, and is available in Theorem III.

By Theorem I itself, there exists a sum $T_{n1}(x)$ such that

$$|f^{(p)}(x) - T_{n1}(x)| \leq \frac{K\lambda}{n}.$$

On the basis of the more recent discussion, as

$$\int_{-\pi}^{\pi} f^{(p)}(x)\, dx = f^{(p-1)}(\pi) - f^{(p-1)}(-\pi) = 0,$$

it may be understood that $T_{n1}(x)$ has no constant term. Since $f^{(p)}(x)$ is the derivative of the periodic function $f^{(p-1)}(x)$, furthermore, an approximating sum $T_{n2}(x)$ may be constructed for $f^{(p-1)}(x)$, as indicated above, with $\varepsilon_n = K\lambda/n$, and

$$|f^{(p-1)}(x) - T_{n2}(x)| \leq \frac{K^2\lambda}{n^2}.$$

If $p > 1$, $f^{(p-1)}(x)$ is itself the derivative of the periodic function $f^{(p-2)}(x)$, and the constant term in $T_{n2}(x)$ is zero. By a sufficient number of repetitions of the process, the theorem is established.

In Theorem II, even if the integral of $f(x)$ over an interval of length 2π is zero, the same thing is not necessarily true of the auxiliary function $\varphi(x)$, and it is not clear that the approximating sum in the conclusion of the theorem will lack the constant term. The difficulty is not a serious one, however. If

$$\int_{-\pi}^{\pi} \varphi(x)\, dx = c,$$

let $\varphi_1(x) = \varphi(x) - c/(2\pi)$. Then

$$\int_{-\pi}^{\pi} \varphi_1(x)\, dx = 0,$$

and there is a trigonometric sum $T_{n1}(x)$, without constant term, such that

$$|\varphi_1(x) - T_{n1}(x)| \leqq \frac{K}{2\pi}\,\omega\left(\frac{2\pi}{n}\right).$$

But on the hypothesis that the integral of $f\,dx$ over a period is zero,

$$|c| = \left|\int_{-\pi}^{\pi} \varphi(x)\,dx\right| = \left|\int_{-\pi}^{\pi} [\varphi(x) - f(x)]\,dx\right|$$
$$\leqq 2\pi \cdot 2\omega\left(\frac{2\pi}{n}\right),$$

and

$$|\varphi(x) - \varphi_1(x)| \leqq 2\omega\left(\frac{2\pi}{n}\right), \quad |f(x) - \varphi_1(x)| \leqq 4\omega\left(\frac{2\pi}{n}\right).$$

So the conclusion of Theorem II applies to the approximate representation of $f(x)$ by a trigonometric sum without constant term, when $f(x)$ is the derivative of a periodic function, on the condition merely that $K' = K/(2\pi)+2$ be replaced by $K'' = K/(2\pi)+4$.

The same process of induction which was used to prove Theorem III then serves to establish

THEOREM IV. *If $f(x)$ is a function of period 2π which has everywhere a continuous pth derivative, with modulus of continuity $\omega(\delta)$, there exists for every positive integral value of n a trigonometric sum $T_n(x)$, of the nth order, such that, for all real values of x,*

$$|f(x) - T_n(x)| \leqq \frac{K'' K^p}{n^p}\,\omega\left(\frac{2\pi}{n}\right),$$

where K is the absolute constant given by the proof of Theorem I, and $K'' = K/(2\pi)+4$.

A part of the content of this theorem may be restated in the following

COROLLARY. *If $f(x)$ is a function of period 2π which has everywhere a continuous pth derivative, there exists for every positive integral value of n a trigonometric sum $T_n(x)$, of the nth order, such that*

$$\lim_{n=\infty} n^p \epsilon_n = 0,$$

if ϵ_n is the maximum of $|f(x) - T_n(x)|$.

2. Approximation by polynomials

The development of the theory of approximation by trigonometric sums will be interrupted at this stage, to make way for a presentation of the corresponding results with regard to polynomial approximation. The transition will be aided, however, by one more lemma on the trigonometric side.

LEMMA. *If $f(x)$ is an even function of period 2π, and if there is a trigonometric sum $T_n(x)$, of the nth order, such that $|f(x) - T_n(x)| \leqq \varepsilon$ for all values of x, there exists a cosine sum $C_n(x)$ of the same order* (that is, a trigonometric sum without sine terms), *such that, for all values of x,*

$$|f(x) - C_n(x)| \leqq \varepsilon.$$

When $f(x)$ is even, the approximating sum given by the proof of Theorem I will automatically lack the sine terms, as an immediate consequence of the fact, already pointed out, that its definition is equivalent to a method of summation of the Fourier series; and this observation would be sufficient for the main argument; but it is of interest to note that the lemma subsists independently of any particular mode of construction of the original approximating function.

For the proof, let

$$C_n(x) = \frac{1}{2}[T_n(x) + T_n(-x)].$$

Then $C_n(x)$ consists merely of the cosine terms of $T_n(x)$, without the sine terms. On the other hand, since $f(x)$ is even,

$$f(x) = \frac{1}{2}[f(x) + f(-x)],$$

and therefore

$$|f(x) - C_n(x)| = \left|\frac{1}{2}[f(x) - T_n(x)] + \frac{1}{2}[f(-x) - T_n(-x)]\right|$$
$$\leqq \varepsilon.$$

Now let $f(x)$ be a function defined for $-1 \leqq x \leqq 1$, and subject to the condition

$$|f(x_2) - f(x_1)| \leqq \lambda |x_2 - x_1|$$

throughout this interval. Let

$$x = \cos\theta, \qquad f(x) = f(\cos\theta) = \varphi(\theta).$$

Then $\varphi(\theta)$ is an even function, defined for all real values of θ, and

$$\begin{aligned}|\varphi(\theta_2) - \varphi(\theta_1)| &= |f(\cos\theta_2) - f(\cos\theta_1)| \\ &\leqq \lambda|\cos\theta_2 - \cos\theta_1| \leqq \lambda|\theta_2 - \theta_1|.\end{aligned}$$

By Theorem I, together with the lemma just proved, there exists a cosine sum $C_n(\theta)$, of the nth order, such that

$$|\varphi(\theta) - C_n(\theta)| \leqq \frac{K\lambda}{n}.$$

But a cosine sum of the nth order in θ is a polynomial of the nth degree in $\cos\theta$, which may be denoted by $P_n(x)$, and the conclusion is that a polynomial $P_n(x)$ exists such that

$$|f(x) - P_n(x)| \leqq \frac{K\lambda}{n}$$

for $-1 \leqq x \leqq 1$.

If the interval $(-1, 1)$ in the hypothesis is replaced by an arbitrary interval (a, b), a preliminary transformation of variable may be made according to the formulas

$$y = \frac{2x - a - b}{b - a}, \qquad f(x) = f_1(y),$$

whereby $f_1(y)$ is defined for $-1 \leqq y \leqq 1$, and

$$\begin{gathered}|f_1(y_2) - f_1(y_1)| \\ = |f(x_2) - f(x_1)| \leqq \lambda|x_2 - x_1| = \frac{1}{2}\lambda(b - a)|y_2 - y_1|.\end{gathered}$$

The result just obtained may then be applied to the approximation of $f_1(y)$ by a polynomial in y, which is at the same time a polynomial in x. The general conclusion may be formulated as

THEOREM V. *If $f(x)$ satisfies the condition*

$$|f(x_2) - f(x_1)| \leqq \lambda|x_2 - x_1|$$

throughout the closed interval (a, b), *of length* l, *there exists for every positive integral value of* n *a polynomial* $P_n(x)$, *of the nth degree, such that*

$$|f(x) - P_n(x)| \leqq \frac{Ll\lambda}{n}$$

for $a \leqq x \leqq b$, *with* $L = \frac{1}{2}K$, *where* K *is the constant of Theorem* I.

If the smallest possible value of K were found in Theorem I, it is not clear that $\frac{1}{2}K$ would then be the *smallest* admissible value of L, but it can be shown that the validity of Theorem V is not general for any L smaller than $\frac{1}{2}$.

More generally still, suppose that $f(x)$ is an arbitrary continuous function for $a \leqq x \leqq b$, and let $\omega(\delta)$ be its modulus of continuity in this interval. With $b - a = l$, let $\varphi(x)$ be the continuous function which takes on the same values as $f(x)$ at the points

$$a,\ a + \frac{l}{n},\ a + \frac{2l}{n},\ \ldots,\ b - \frac{l}{n},\ b,$$

and is linear from each point of this set to the next. The function $\varphi(x)$, having a broken line for its graph, satisfies the hypothesis of Theorem V, with

$$\lambda = \frac{\omega(l/n)}{l/n}.$$

while

$$|f(x) - \varphi(x)| \leqq 2\omega(l/n)$$

throughout (a, b). There is a polynomial $P_n(x)$ such that

$$|\varphi(x) - P_n(x)| \leqq L\omega(l/n);$$

setting $L + 2 = L'$, one may state

THEOREM VI. *If* $f(x)$ *is a continuous function with modulus of continuity* $\omega(\delta)$ *in the closed interval* (a, b), *of length* l, *there exists for every positive integral value of* n *a polynomial* $P_n(x)$, *of the nth degree, such that, for* $a \leqq x \leqq b$,

$$|f(x) - P_n(x)| \leqq L'\omega(l/n),$$

where L' *is an absolute constant.*

This incidentally includes the theorem of Weierstrass on polynomial approximation, which was quoted at the beginning of the chapter.

The proof of a theorem corresponding to Theorem III is simplified by the fact that the indefinite integral of a polynomial is always a polynomial, so that special considerations analogous to those relating to the constant term in the trigonometric case are not needed. A new complication is introduced, on the other hand, by the circumstance that the degree of a polynomial is raised by integration, while the order of a trigonometric sum remains unchanged.

Suppose that $f(x)$ has a continuous derivative $f'(x)$ for $a \leq x \leq b$, and that there is a polynomial $p'_n(x)$, *of degree* $n-1$, such that

$$|f'(x) - p'_n(x)| \leq \epsilon_n$$

throughout the interval. Let

$$\int_a^x p'_n(x)\,dx = p_n(x), \qquad f(x) - p_n(x) = r_n(x).$$

Since $|r'_n(x)| \leq \epsilon_n$, $r_n(x)$ satisfies the hypothesis of Theorem V, with $\lambda = \epsilon_n$. There is consequently a polynomial $\pi_n(x)$, of the nth degree, such that

$$|r_n(x) - \pi_n(x)| \leq \frac{Ll\epsilon_n}{n}.$$

If $p_n(x) + \pi_n(x) = P_n(x)$, this $P_n(x)$ is a polynomial of the nth degree, and

$$|f(x) - P_n(x)| = |r_n(x) - \pi_n(x)| \leq \frac{Ll\epsilon_n}{n}.$$

Let $f(x)$ have a pth derivative $f^{(p)}(x)$, satisfying the condition that

$$|f^{(p)}(x_2) - f^{(p)}(x_1)| \leq \lambda |x_2 - x_1|$$

throughout (a, b). If $n - p > 0$, there is a polynomial of degree $n-p$ which differs from $f^{(p)}(x)$ by not more than $Ll\lambda/(n-p)$ throughout the interval. There is then, by the preceding paragraph, a polynomial of degree $n-p+1$,

differing by not more than $L^2 l^2 \lambda / [(n-p)(n-p+1)]$ from $f^{(p-1)}(x)$, and so on. Finally a polynomial $P_n(x)$ is obtained, of the nth degree, for which

$$|f(x) - P_n(x)| \leqq \frac{L^{p+1} l^{p+1} \lambda}{(n-p)(n-p+1) \cdots n}.$$

For the applications, the existence of a constant in the right-hand member of the last relation is far more important than any close estimate of its numerical value. It is most convenient, even at considerable unnecessary expense numerically, to be satisfied with the observation that

$$\frac{1}{(n-p)(n-p+1)\cdots n} = \frac{n}{n-p} \cdot \frac{n}{n-p+1} \cdots$$

$$\cdots \frac{n}{n-1} \cdot \frac{1}{n^{p+1}} \leqq \frac{(p+1)^p}{p!} \cdot \frac{1}{n^{p+1}}$$

for $n \geqq p+1$, and to state the result in the form of

THEOREM VII. *If $f(x)$ has a pth derivative $f^{(p)}(x)$ satisfying the condition that*

$$|f^{(p)}(x_2) - f^{(p)}(x_1)| \leqq \lambda |x_2 - x_1|$$

throughout the closed interval (a, b), of length l, there exists for every integral value of $n > p$ a polynomial $P_n(x)$, of the nth degree, such that for $a \leqq x \leqq b$,

$$|f(x) - P_n(x)| \leqq \frac{L_p l^{p+1} \lambda}{n^{p+1}},$$

where $L_p = (p+1)^p L^{p+1}/p!$, and L is the constant of Theorem V.

It is clear that even with $\lambda = 0$ the hypothesis implies nothing whatever as to the possibility of approximating $f(x)$ by a polynomial of degree lower than p, since $f(x)$ itself may then be any polynomial of the pth degree. By suitable changes in formulation it would be possible, though of secondary interest, to admit the value $n = 0$ in Theorem V (or Theorem I), and the value $n = p$ here.

From Theorem VI, by reasoning similar to the above, one may deduce

THEOREM VIII. *If $f(x)$ has a continuous pth derivative with modulus of continuity $\omega(\delta)$ throughout the closed interval (a, b), of length l, there exists for every integral value of $n > p$ a polynomial $P_n(x)$, of the nth degree, such that for $a \leqq x \leqq b$,*

$$|f(x) - P_n(x)| \leqq \frac{L'_p l^p}{n^p} \omega\left(\frac{l}{n-p}\right),$$

where $L'_p = (p+1)^{p-1} L^p (L+2)/p!$, and L is the constant of Theorem V.

COROLLARY. *If $f(x)$ has a continuous pth derivative for $a \leqq x \leqq b$, there exists for every positive integral value of n a polynomial $P_n(x)$, of the nth degree, such that*

$$\lim_{n=\infty} n^p \varepsilon_n = 0,$$

where ε_n is the maximum of $|f(x) - P_n(x)|$ in the interval (a, b).

The exceptional status of the values of $n \leqq p$ has no significance for the corollary, which is concerned only with a limit for $n = \infty$.

3. Degree of convergence of Fourier series

The preceding theorems can be made to serve as basis for a discussion of the convergence and rapidity of convergence of Fourier and Legendre series, as well as of other processes of approximation. This idea will be developed more fully in succeeding chapters; its first consequences will be presented here.

With regard to the Fourier series for a given function $f(x)$, it will be premised merely that it is a series of the form

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx),$$

in which the coefficients have the values

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos kt\, dt, \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin kt\, dt,$$

the formula for a_k being applicable when $k = 0$, as well as when k is positive. The expression

$$S_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

will be called the "partial sum of the series to terms of the nth order."

If the quantity

$$\frac{1}{2} + \cos u + \cos 2u + \cdots + \cos nu$$

is multiplied by $2 \sin \frac{1}{2} u$, the product may be rewritten as

$$\sin \frac{u}{2} + \left[\sin \frac{3u}{2} - \sin \frac{u}{2}\right] + \cdots$$
$$\cdots + \left[\sin \left(n + \frac{1}{2}\right) u - \sin \left(n - \frac{1}{2}\right) u\right],$$

which immediately reduces to $\sin (n + \frac{1}{2}) u$, so that

$$\frac{1}{2} + \cos u + \cos 2u + \cdots + \cos nu = \frac{\sin (n + \frac{1}{2}) u}{2 \sin \frac{1}{2} u}.$$

Let each coefficient in $S_n(x)$ be replaced by the integral expression which defines it. Since $\cos kx$ and $\sin kx$ are independent of t, they may be written under the sign of integration, and the various integrals may be combined into a single one:

$$\begin{aligned} S_n(x) &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{k=1}^{n} (\cos kt \cos kx + \sin kt \sin kx)\right] dt \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{k=1}^{n} \cos k(t - x)\right] dt \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \frac{\sin (n + \frac{1}{2})(t - x)}{2 \sin \frac{1}{2}(t - x)} dt. \end{aligned}$$

From the last expression may be deduced the following:

LEMMA. *If $f(x)$, of period 2π, is bounded and integrable* (in the sense of Riemann or in the sense of Lebesgue), *if*

$$|f(x)| \leq M$$

for all values of x, and if $S_n(x)$ is the partial sum of the Fourier series for $f(x)$, to terms of the nth order, then

$$|S_n(x)| \leqq CM \log n,$$

for all values of x and for all values of $n > 1$, where C is an absolute constant, depending neither on x, nor on n, nor on the function $f(x)$.

By the hypothesis on $f(x)$,

$$|S_n(x)| \leqq \frac{M}{\pi}\int_{-\pi}^{\pi}\left|\frac{\sin(n+\frac{1}{2})(t-x)}{2\sin\frac{1}{2}(t-x)}\right| dt.$$

If one-half the integral on the right is denoted by j_n, its form may be modified by the substitution $u = \frac{1}{2}(t-x)$, and by recognition of the fact that the resulting integrand is an even function of u, of period π, to yield the conclusion that

$$j_n = \frac{1}{2}\int_{-(\pi+x)/2}^{(\pi-x)/2}\left|\frac{\sin(2n+1)u}{\sin u}\right| du$$

$$= \frac{1}{2}\int_{-\pi/2}^{\pi/2}\left|\frac{\sin(2n+1)u}{\sin u}\right| du = \int_0^{\pi/2}\left|\frac{\sin(2n+1)u}{\sin u}\right| du,$$

$$|S_n(x)| \leqq \frac{2M}{\pi} j_n.$$

From the fact that

$$\frac{\sin(2n+1)u}{\sin u} = 1 + 2\sum_{k=1}^{n}\cos 2ku$$

it follows that

$$\left|\frac{\sin(2n+1)u}{\sin u}\right| \leqq 2n+1$$

for all values of u. On the other hand, $|\sin(2n+1)u| \leqq 1$, while

$$\frac{\sin u}{u} \geqq \frac{\sin(\pi/2)}{\pi/2} = \frac{2}{\pi}, \qquad \frac{1}{\sin u} \leqq \frac{\pi}{2}\cdot\frac{1}{u}$$

throughout the interval of integration. So

$$j_n = \int_0^{1/n} + \int_{1/n}^{\pi/2} \leqq \int_0^{1/n}(2n+1)\,du + \frac{\pi}{2}\int_{1/n}^{\pi/2}\frac{du}{u}$$

$$= 2 + \frac{1}{n} + \frac{\pi}{2}\log\frac{\pi}{2} + \frac{\pi}{2}\log n.$$

The last expression does not exceed a constant multiple of $\log n$, for $n \geqq 2$, and the conclusion of the lemma is justified.

The method of application of the lemma, which is due to Lebesgue (though he did not make so extensive use of it as is done here), may be summarized in

THEOREM IX. *If $f(x)$ is a continuous function of period 2π, and $S_n(x)$ the partial sum of its Fourier series to terms of the nth order, $n \geqq 1$, and if there exists a trigonometric sum $T_n(x)$, of the nth order, such that*

$$|f(x) - T_n(x)| \leqq \epsilon_n$$

for all values of x, then, for all values of x,

$$|f(x) - S_n(x)| \leqq B\epsilon_n \log n,$$

where B is an absolute constant.

The statement is equally true, though of less interest, if $f(x)$ is merely assumed to be integrable; of course ϵ_n can not approach zero, when the relations are considered for successive values of n, unless $f(x)$ is continuous.

Let the sum $T_n(x)$ in the hypothesis have the expression

$$\frac{\alpha_0}{2} + \sum_{k=1}^{n} (\alpha_k \cos kx + \beta_k \sin kx).$$

It is found by direct integration that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} T_n(t) \cos kt\, dt = \alpha_k, \qquad \frac{1}{\pi}\int_{-\pi}^{\pi} T_n(t) \sin kt\, dt = \beta_k,$$

for $k \leqq n$, so that *the partial sum of the Fourier series for $T_n(x)$, to terms of order n, is identical with $T_n(x)$ itself.* If the Fourier coefficients of $f(x)$ are a_k, b_k, those of the function

$$R_n(x) = f(x) - T_n(x)$$

are $a_k - \alpha_k$, $b_k - \beta_k$, for $k \leqq n$, and the partial sum $s_n(x)$ of the Fourier series for $R_n(x)$, to terms of order n, is

$$\begin{aligned} s_n(x) &= \frac{1}{2}(a_0 - \alpha_0) + \sum_{k=1}^{n} [(a_k - \alpha_k) \cos kx + (b_k - \beta_k) \sin kx] \\ &= S_n(x) - T_n(x). \end{aligned}$$

Since $|R_n(x)| \leqq \varepsilon_n$, it follows from the Lemma that

$$|s_n(x)| \leqq C\varepsilon_n \log n.$$

Consequently

$$|f(x) - S_n(x)| \leqq |f(x) - T_n(x)| + |T_n(x) - S_n(x)|$$
$$= |R_n(x)| + |s_n(x)| \leqq \varepsilon_n + C\varepsilon_n \log n \leqq \varepsilon_n \frac{\log n}{\log 2} + C\varepsilon_n \log n$$

for $n \geqq 2$, and the last expression has the form $B\varepsilon_n \log n$, with $B = (\log 2)^{-1} + C$.

The theorem may immediately be specialized and made more definite by combination with Theorems I—IV, as follows:

COROLLARY I. *If*

$$|f(x_2) - f(x_1)| \leqq \lambda |x_2 - x_1|$$

for all values of x_1 *and* x_2, λ *being a constant, then*

$$|f(x) - S_n(x)| \leqq \frac{A\lambda \log n}{n}.$$

COROLLARY II. *If* $f(x)$ *is continuous with modulus of continuity* $\omega(\delta)$,

$$|f(x) - S_n(x)| \leqq A\omega\left(\frac{2\pi}{n}\right) \log n.$$

COROLLARY IIa. *The Fourier series converges uniformly to the value* $f(x)$, *if* $f(x)$ *has a modulus of continuity* $\omega(\delta)$ *such that* $\lim_{\delta=0} \omega(\delta) \log \delta = 0$ (Lipschitz-Dini condition).

COROLLARY III. *If* $f(x)$ *has a pth derivative* $f^{(p)}(x)$ *such that*

$$|f^{(p)}(x_2) - f^{(p)}(x_1)| \leqq \lambda |x_2 - x_1|$$

for all values of x_1 *and* x_2, λ *being a constant, then*

$$|f(x) - S_n(x)| \leqq \frac{A_p \lambda \log n}{n^{p+1}}.$$

COROLLARY IV. *If* $f(x)$ *has everywhere a continuous pth derivative with modulus of continuity* $\omega(\delta)$,

$$|f(x) - S_n(x)| \leqq \frac{A_p}{n^p} \omega\left(\frac{2\pi}{n}\right) \log n.$$

In each of these statements, the conclusion holds for all values of x, and for all values of $n \geqq 2$; the coefficient A is an absolute constant, while A_p depends only on p.

The corollaries are stated separately for emphasis; it is clear that all are included in Corollary IV, if it is understood that p may in particular have the value 0. The use of the same A in I and II, and of the same A_p in III and IV, signifies merely that when two constants are concerned, one symbol may be used to represent the larger of them.

It is worthy of note that even the dependence of A_p on p can be eliminated, with a resulting simplification in Corollaries III and IV which does not have a counterpart, as far as the present evidence goes, in the case of the corresponding third and fourth Theorems.

If $f(x)$ is a function of period 2π having a continuous derivative, the Fourier series for $f'(x)$ is that obtained by formal differentiation of the Fourier series for $f(x)$. This is recognized without any further assumption as to the convergence of the series, from the relations

$$\int_{-\pi}^{\pi} f'(t) \cos k t \, dt = k \int_{-\pi}^{\pi} f(t) \sin k t \, dt,$$
$$\int_{-\pi}^{\pi} f'(t) \sin k t \, dt = -k \int_{-\pi}^{\pi} f(t) \cos k t \, dt,$$

which are obtained by integration by parts, the terms which would appear outside the integral sign reducing to zero, because of the periodicity of the functions involved. If the coefficients in the series for $f(x)$ once more are a_k, b_k, and if $f(x)$ has a continuous derivative of order $2q$, where q is any positive integer, the Fourier coefficients for $f^{(2q)}(x)$ are $\alpha_k = (-1)^q k^{2q} a_k$, $\beta_k = (-1)^q k^{2q} b_k$, which means that the series for $f(x)$ can be written in the form

$$\frac{a_0}{2} + (-1)^q \sum_{k=1}^{\infty} \frac{1}{k^{2q}} (\alpha_k \cos kx + \beta_k \sin kx).$$

The partial sums of the series for $f(x)$ and for $f^{(2q)}(x)$, to terms of the nth order, may be represented by $S_n(x)$ and $S_n^{(2q)}(x)$ respectively.

Let

$$f^{(2q)}(x) - S_n^{(2q)}(x) = \varrho_n(x),$$

and suppose now that $|\varrho_n(x)| \leqq \varepsilon_n$, where $\varepsilon_{n+1} \leqq \varepsilon_n$ for all values of n that are considered, and $\lim_{n=\infty} \varepsilon_n = 0$. Then

$$\alpha_k \cos kx + \beta_k \sin kx = \varrho_{k-1}(x) - \varrho_k(x).$$

By any of the preceding corollaries, IIa, for example, the series for $f(x)$ converges to the value of the function, while the series $\sum \varrho_{k-1}(x)/k^{2q}$, $\sum \varrho_k(x)/k^{2q}$ are convergent, because $|\varrho_k(x)|$ is uniformly bounded. Hence the remainder $f(x) - S_n(x)$ can be written and rearranged as follows:

$$\begin{aligned} f(x) - S_n(x) &= (-1)^q \sum_{k=n+1}^{\infty} \frac{1}{k^{2q}} (\alpha_k \cos kx + \beta_k \sin kx) \\ &= (-1)^q \sum_{k=n+1}^{\infty} \frac{1}{k^{2q}} [\varrho_{k-1}(x) - \varrho_k(x)] \\ &= (-1)^{q+1} \left[-\frac{1}{(n+1)^{2q}} \varrho_n(x) \right. \\ &\qquad \left. + \sum_{k=n+1}^{\infty} \left(\frac{1}{k^{2q}} - \frac{1}{(k+1)^{2q}} \right) \varrho_k(x) \right]. \end{aligned}$$

As the parentheses in the last summation are all positive, and as $|\varrho_k(x)| \leqq \varepsilon_k \leqq \varepsilon_n$ for $k > n$,

$$\begin{aligned} |f(x) - S_n(x)| &\leqq \frac{\varepsilon_n}{(n+1)^{2q}} + \varepsilon_n \sum_{k=n+1}^{\infty} \left(\frac{1}{k^{2q}} - \frac{1}{(k+1)^{2q}} \right) \\ &= \frac{2\varepsilon_n}{(n+1)^{2q}} < \frac{2\varepsilon_n}{n^{2q}}. \end{aligned}$$

Under the hypotheses of Corollary III, let $p = 2q$ if p is even, $p = 2q+1$ if p is odd. Then Corollary I or Corollary III may be applied directly to $f^{(2q)}(x)$, which is found to satisfy the requirements of the preceding paragraph, with $\varepsilon_n = (A\lambda \log n)/n$ in one case and $\varepsilon_n = (A_1 \lambda \log n)/n^2$ in the other. The corresponding values of $2\varepsilon_n/n^{2q}$ are $(2A\lambda \log n)/n^{p+1}$ and $(2A_1\lambda \log n)/n^{p+1}$ respectively. For literal accuracy, it must be admitted that $(\log n)/n$ diminishes only from $n = 3$ on, and that the conclusion has been established, when p is even,

only for $n \geqq 3$. The value $n = 2$ may be included if the previous ϵ_2 is replaced by a somewhat larger quantity, with a suitable adjustment of the constant in the conclusion, if necessary.

Similar reasoning may be employed under the conditions of Corollary IV, except that it is no longer satisfactory to take $2q = p$ when p is even, as $\omega(2\pi/n)\log n$ might not decrease with increasing n. It is sufficient, however, to let $p = 2q+2$ when p is even, $p = 2q+1$ when p is odd, and to obtain ϵ_n by the application of Corollary IV to $f^{(2q)}(x)$ as thus defined. From the definition of $\omega(\delta)$ it is certain that $\omega(2\pi/n)$ itself diminishes, or at any rate does not increase, as n increases.

All the cases in question are covered, with some redundancy, if the following are considered successively: for $\omega(\delta) \leqq \lambda\delta$, $p = 0, 1, 2q \ (q \geqq 1), 2q+1 \ (q \geqq 1)$; for general $\omega(\delta)$, $p = 0, 1, 2, 2q+1 \ (q \geqq 1), 2q+2 \ (q \geqq 1)$. If a single letter is used to represent the largest of the finite number of constants entering into the corresponding conclusions, the result may be formulated thus:

THEOREM X. *The preceding Corollaries* III *and* IV *may be restated, for all values of $p \geqq 0$, with the multiplier A_p replaced by an absolute constant D, depending neither on p nor on anything else.*

4. Degree of convergence of Legendre series

A considerable part of the above reasoning can be carried over to the case of Legendre series, though the relations are less simple than for Fourier series, and the results as presented here will be less complete.

By the Legendre series for a given continuous function $f(x)$ is meant a series of the form

$$a_0 X_0(x) + a_1 X_1(x) + a_2 X_2(x) + \cdots$$

where $X_k(x)$ is the Legendre polynomial of the kth degree, and

$$a_k = \frac{2k+1}{2} \int_{-1}^{1} f(t)\, X_k(t)\, dt.$$

Not to work out the theory of these polynomials from the beginning, it will be assumed as known that one of them is defined for each integral value of $k \geqq 0$, the first two being $X_0 = 1$, $X_1 = x$; that they satisfy the relations

$$\int_{-1}^{1} X_j(t)\, X_k(t)\, dt = 0 \quad (j \neq k), \qquad \int_{-1}^{1} X_k^2(t)\, dt = \frac{2}{2k+1};$$

that any successive three of them are connected by the recursion formula

$$(k+1)\, X_{k+1}(x) - (2k+1)\, x\, X_k(x) + k X_{k-1}(x) = 0;$$

and that the polynomial of the kth degree can be expressed in the form

$$X_k(x) = \frac{1}{\pi} \int_0^{\pi} [x + i(1-x^2)^{1/2} \cos\varphi]^k \, d\varphi .$$

In the last expression, the presence of imaginaries is only superficially apparent; if the integrand is expanded by the binomial theorem for a positive integral exponent, the coefficient of each odd power of i after integration is an integral which is seen at once to be equal to zero.

Let $S_n(x)$ stand for the sum of the first $n+1$ terms of the series:

$$S_n(x) = a_0 X_0(x) + a_1 X_1(x) + \cdots + a_n X_n(x).$$

By the definition of the coefficients,

$$S_n(x) = \int_{-1}^{1} f(t)\, K_n(x, t)\, dt,$$

where

$$K_n(x, t) = \frac{1}{2} [X_0(x)\, X_0(t) + 3 X_1(x)\, X_1(t) + \cdots$$
$$\cdots + (2n+1)\, X_n(x)\, X_n(t)].$$

This function can be rewritten in the form

$$K_n(x, t) = \frac{n+1}{2} \cdot \frac{X_{n+1}(x)\, X_n(t) - X_n(x)\, X_{n+1}(t)}{x - t};$$

the identity is immediately verified for $n = 0$, and may then be proved in general by a straightforward process of induction. based on the recursion formula.

In the integral expression for $X_k(x)$, when $x \leq 1$,

$$|x+i(1-x^2)^{1/2}\cos\varphi| = [x^2+(1-x^2)\cos^2\varphi]^{1/2}$$
$$\leq [x^2+(1-x^2)]^{1/2} = 1,$$

and hence $|X_k(x)| \leq 1$. There will be occasion to use a closer inequality for $|X_k|$ in the interior of the interval $(-1, 1)$. To return to the integral representation,

$$|X_k(x)| \leq \frac{1}{\pi}\int_0^{\pi}[x^2+(1-x^2)\cos^2\varphi]^{k/2}\,d\varphi$$
$$= \frac{2}{\pi}\int_0^{\pi/2}[x^2+(1-x^2)\cos^2\varphi]^{k/2}\,d\varphi,$$

the last equality resulting from the fact that $\cos^2\varphi = \cos^2(\pi-\varphi)$. Since

$$x^2+(1-x^2)\cos^2\varphi = \cos^2\varphi+x^2\sin^2\varphi = 1-(1-x^2)\sin^2\varphi,$$

and since $\sin\varphi \geq 2\varphi/\pi$ throughout the interval of integration, while $1-x^2$ is positive for the values of x under consideration, it follows that

$$x^2+(1-x^2)\cos^2\varphi \leq 1-\frac{4}{\pi^2}(1-x^2)\varphi^2 = 1-\xi^2\varphi^2,$$

if $\xi = 2(1-x^2)^{1/2}/\pi$. By an application of the extended mean value theorem to the function e^{-y},

$$e^{-y} = 1-y+\frac{1}{2}y^2e^{-\theta y}, \quad 0<\theta<1,$$

so that $e^{-y} \geq 1-y$ for all real values of y, and in the present connection

$$1-\xi^2\varphi^2 \leq e^{-\xi^2\varphi^2}.$$

So

$$|X_k(x)| \leq \frac{2}{\pi}\int_0^{\pi/2} e^{-k\xi^2\varphi^2/2}\,d\varphi \leq \frac{2}{\pi}\int_0^{\infty} e^{-k\xi^2\varphi^2/2}\,d\varphi,$$

and, by the substitution $u = (\frac{1}{2}k)^{1/2}\xi\varphi$,

$$|X_k(x)| \leq \frac{2^{3/2}}{\pi\xi k^{1/2}}\int_0^\infty e^{-u^2}\,du = \frac{c_1}{k^{1/2}(1-x^2)^{1/2}} \qquad (k > 0),$$

where c_1 is independent of k and x; the numerical value $c_1 = (\pi/2)^{1/2}$ is not essential for present purposes.

From the relation just obtained it follows, first, that if x is restricted to an interval $-1+\frac{1}{2}\eta \leq x \leq 1-\frac{1}{2}\eta$, $0 < \eta < 2$, then

$$|X_k(x)| \leq \frac{g}{k^{1/2}},$$

where g is independent of k and x, but depends on η; and secondly, that

$$\int_{-1}^{1}|X_k(x)|\,dx \leq \frac{c_1}{k^{1/2}}\int_{-1}^{1}\frac{dx}{(1-x^2)^{1/2}} = \frac{c_2}{k^{1/2}},$$

where c_2 is independent of k.

It is possible now to proceed to the proof of the following lemma, which assigns an upper bound for $|S_n(x)|$, not throughout the entire interval $-1 \leq x \leq 1$, to be sure, but throughout an interval interior to it:

LEMMA. *If $f(x)$ is bounded and integrable* (in the sense of Riemann or in the sense of Lebesgue) *for* $-1 \leq x \leq 1$, *if*

$$|f(x)| \leq M$$

throughout the interval, and if $S_n(x)$ is the partial sum of the Legendre series for $f(x)$, then

$$|S_n(x)| \leq GM\log n$$

for $-1+\eta \leq x \leq 1-\eta$, $0 < \eta < 1$, where G does not depend on x, n, or the function $f(x)$, but does depend on η.

The proof starts from the fact that

$$|S_n(x)| \leq M\int_{-1}^{1}|K_n(x,t)|\,dt.$$

Let the interval of integration be divided into five sub-intervals, terminated by the points

$$-1,\ -1+\frac{1}{2}\eta,\ x-\frac{1}{n},\ x+\frac{1}{n},\ 1-\frac{1}{2}\eta,\ 1.$$

Under the hypothesis that $-1+\eta \leqq x \leqq 1-\eta$, these points will succeed each other in the order named, as soon as $n > 2/\eta$. It will be assumed for the present that this condition is satisfied; then, in particular, $n \neq 0$, and as there will be occasion to observe incidentally, $[(n+1)/n]^{1/2} < 2$. Let the values of the integral of $|K_n|$ over the sub-intervals be denoted by $I_1, \cdots, I_5$ respectively. The relation $|X_k(t)| \leqq g/k^{1/2}$ can be used in the second, third and fourth integrals, and the relation $|X_k(x)| \leqq g/k^{1/2}$, which is independent of t, in any of the five. In the middle interval,

$$
\begin{aligned}
|K_n(x, t)| &= \left|\sum_{k=0}^{n} \frac{1}{2}(2k+1) X_k(x) X_k(t)\right| \\
&\leqq \frac{1}{2} + \sum_{k=1}^{n} \frac{1}{2}(2k+1) \cdot \frac{g}{k^{1/2}} \cdot \frac{g}{k^{1/2}} \\
&= \frac{1}{2} + \sum_{k=1}^{n} g^2\left(1+\frac{1}{2k}\right) \\
&\leqq 1 + \sum_{k=1}^{n} 2g^2 = 1 + 2g^2 n \leqq (2g^2+1) n,
\end{aligned}
$$

and hence

$$
I_3 = \int_{x-1/n}^{x+1/n} |K_n(x, t)|\, dt \leqq \int_{x-1/n}^{x+1/n} (2g^2+1) n\, dt = 2(2g^2+1).
$$

The representation of $K_n(x, t)$ as a fraction is to be used in the remaining integrals. In the first interval, $|x-t| \geqq \frac{1}{2}\eta$, and

$$
\begin{aligned}
I_1 &= \int_{-1}^{-1+\frac{1}{2}\eta} |K_n(x, t)|\, dt \\
&\leqq \frac{n+1}{\eta}\left[|X_{n+1}(x)| \int_{-1}^{-1+\frac{1}{2}\eta} |X_n(t)|\, dt\right. \\
&\qquad\qquad \left. + |X_n(x)| \int_{-1}^{-1+\frac{1}{2}\eta} |X_{n+1}(t)|\, dt\right] \\
&\leqq \frac{n+1}{\eta}\left[\frac{g}{(n+1)^{1/2}} \int_{-1}^{1} |X_n(t)|\, dt + \frac{g}{n^{1/2}} \int_{-1}^{1} |X_{n+1}(t)|\, dt\right] \\
&\leqq \frac{n+1}{\eta} \cdot \frac{2c_2 g}{[n(n+1)]^{1/2}} < \frac{4c_2 g}{\eta}.
\end{aligned}
$$

In the same way, $I_5 < 4c_2 g/\eta$. In the second interval,

$$\frac{n+1}{2} \,|\, X_{n+1}(x) X_n(t) - X_n(x) X_{n+1}(t) \,|$$
$$\leqq \frac{n+1}{2} \cdot \frac{2g^2}{[n(n+1)]^{1/2}} < 2g^2,$$

so that

$$I_2 = \int_{-1+\frac{1}{2}\eta}^{x-1/n} |K_n(x, t)|\, dt < 2g^2 \int_{-1+\frac{1}{2}\eta}^{x-1/n} \frac{dt}{x-t},$$

or, by the substitution $x - t = u$,

$$I_2 < 2g^2 \int_{1/n}^{x+1-\frac{1}{2}\eta} \frac{du}{u} < 2g^2 \int_{1/n}^{2} \frac{du}{u} = 2g^2 (\log 2 + \log n).$$

Similarly, $I_4 < 2g^2(\log 2 + \log n)$. As n is an integer satisfying the condition $n > 2/\eta > 2$, it is certain that $n \geqq 3$. $\log n > 1$, and the inequalities that have been obtained for $I_1, \cdots, I_5$ will merely be strengthened if the factor $\log n$ is inserted in the right-hand members wherever it does not occur. By combination of these inequalities,

$$\int_{-1}^{1} |K_n(x, t)|\, dt \leqq G_1 \log n$$

for $n > 2/\eta$, the number G_1 depending only on η. For each value of n belonging to the range $2 \leqq n \leqq 2/\eta$, the integral, considered as a function of x for $-1+\eta \leqq x \leqq 1-\eta$, has a maximum value. Let G_2 be the largest of the finite number of maxima thus determined. Then G_2 depends only on η, and the statement of the lemma is true for all values of $n > 2$, if G is taken as the larger of the numbers G_1, $G_2/\log 2$.

For the application of the lemma, it is to be noticed that an arbitrary polynomial of the nth degree can be expressed identically as a linear combination of $X_0(x), \cdots, X_n(x)$, with constant coefficients. Then a process of reasoning which is entirely analogous to that used in the proof of Theorem IX, and which need not be repeated at length, serves to demonstrate

THEOREM XI. *If $f(x)$ is a continuous function for $-1 \leqq x \leqq 1$, and $S_n(x)$ the sum of the first $n+1$ terms of its Legendre series, $n > 1$, and if there exists a polynomial $P_n(x)$, of the nth degree, such that*

$$|f(x) - P_n(x)| \leqq \epsilon_n$$

for $-1 \leqq x \leqq 1$, then, for $-1+\eta \leqq x \leqq 1-\eta$ $(\eta > 0)$,

$$|f(x) - S_n(x)| \leqq H\epsilon_n \log n,$$

where H depends only on η.

The more specific results obtained by combining this proposition with Theorems V—VIII will not be formulated separately, but will be summarized in a single

COROLLARY. *If $f(x)$ has a continuous pth derivative $(p \geqq 0)$ for $-1 \leqq x \leqq 1$, with modulus of continuity $\omega(\delta)$, and if $S_n(x)$ is the sum of the first $n+1$ terms of the Legendre series for $f(x)$, then*

$$|f(x) - S_n(x)| \leqq \frac{H_p}{n^p}\, \omega\left(\frac{2}{n-p}\right) \log n,$$

for $n > p$ and for $-1+\eta \leqq x \leqq 1-\eta$, where H_p depends on η and on p, but not on anything else: in particular, the series converges uniformly to the value $f(x)$ for $-1+\eta \leqq x \leqq 1-\eta$, if $f(x)$ itself has a modulus of continuity $\omega(\delta)$ for $-1 \leqq x \leqq 1$, such that $\lim_{\delta=0} \omega(\delta) \log \delta = 0$.

Since η may be arbitrarily small, the last conclusion implies that the series is convergent, not necessarily uniformly, throughout the open interval $-1 < x < 1$.

Gronwall (Mathematische Annalen, vol. 74 (1913), pp. 213–270; Transactions of the American Mathematical Society, vol. 15 (1914), pp. 1–30) has shown essentially that

$$\int_{-1}^{1} |K_n(x, t)|\, dt \leqq G_0\, n^{1/2}$$

for $-1 \leqq x \leqq 1$, where G_0 is an absolute constant. It follows that conclusions analogous to those of the preceding theorem and its corollary hold for the entire closed interval $(-1, 1)$,

if the factor $\log n$ is replaced by $n^{1/2}$, the corresponding sufficient condition for uniform convergence being that $\lim_{\delta=0} \omega(\delta)/\delta^{1/2} = 0$. But the proof of the relation of inequality for the integral appears to be rather long, and will not be set forth here. From the definition of K_n, together with the facts that $|X_k(x)| \leq 1$ for $-1 \leq x \leq 1$ and that $\int X_k(t)\,dt \leq c_2/k^{1/2}$, it is obvious that

$$\int_{-1}^{1} |K_n(x, t)|\,dt \leq 1 + \frac{1}{2}\sum_{k=1}^{n} (2k+1)\,c_2/k^{1/2} < c_3\,n^{3/2},$$

where c_3 is an absolute constant; and it can be inferred at once that the series converges uniformly to the value $f(x)$ for $-1 \leq x \leq 1$, if $f(x)$ has a first derivative with a modulus of continuity $\omega(\delta)$ such that $\lim_{\delta=0} \omega(\delta)/\delta^{1/2} = 0$, while there are corresponding theorems on degree of convergence. In a later chapter, conditions will be obtained which are closer than those thus indicated, though not so close as the ones corresponding to the factor $n^{1/2}$.

From the discussion of approximation in terms of polynomials and trigonometric sums it is natural to pass to similar questions with regard to developments in series of more general functions, particularly the characteristic functions defined by linear differential equations with boundary conditions. The beginnings of such a theory have been presented by the author (Transactions of the American Mathematical Society, vol. 15 (1914), pp. 439–466) and W. E. Milne (the same Transactions, vol. 19 (1918), pp. 143–156). The present account, however, will be continued along other lines.

CHAPTER II

DISCONTINUOUS FUNCTIONS; FUNCTIONS OF LIMITED VARIATION; ARITHMETIC MEANS

Introduction

The discussion hitherto has been concerned almost entirely with uniform convergence, and with functions that are continuous throughout the interval under consideration. Corresponding theorems with regard to the approximate representation of discontinuous functions are naturally less simple, and perhaps of less immediate interest. It will be well, nevertheless, not to disregard entirely the question how far the theory that has been outlined can be brought to bear on the representation of such functions. This question will occupy the present chapter. (For a somewhat different set of theorems on the approximate representation of discontinuous functions, reference may be made to a paper by C. E. Wilder, in the Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), pp. 345–361.) There will be occasion also for the further development of the theory as applied to continuous functions, by reference to the concept of limited variation, and by study of the summation of Fourier series according to the method of the first arithmetic mean. A beginning will be made by a review of some well-known general theorems about Fourier series.

1. Convergence of Fourier series under hypothesis of continuity over part of a period

Let $f(x)$ be a function of period 2π, which is summable over a period, together with its square. (Without change in the form of the argument, the discussion can be kept elementary for the present, with results which still have a high degree

of generality, if it is assumed instead that $f(x)$ is bounded and integrable in the sense of Riemann.) Let

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos kt\,dt, \quad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin kt\,dt,$$

$$S_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n}(a_k\cos kx + b_k\sin kx).$$

Then, as may be verified by multiplying out the square and integrating term by term,

$$\frac{1}{\pi}\int_{-\pi}^{\pi}[f(x) - S_n(x)]^2\,dx$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x)]^2\,dx - \frac{2}{\pi}\int_{-\pi}^{\pi}f(x)S_n(x)\,dx + \frac{1}{\pi}\int_{-\pi}^{\pi}[S_n(x)]^2\,dx$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x)]^2\,dx - 2\left[\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right]$$

$$+ \left[\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right]$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x)]^2\,dx - \left[\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right].$$

As the first member can not be negative, it must be that

$$\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2) \leqq \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x)]^2\,dx$$

for all values of n, and hence that the series

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2)$$

is convergent. It follows that for any function $f(x)$ of the character specified, the coefficients a_k, b_k approach zero as k becomes infinite. The identification of the sum of the series with the value $(1/\pi)\int [f(x)]^2\,dx$ is not needed for present purposes.

The assumption that $[f(x)]^2$ is summable is not essential to the truth of the conclusion. On the hypothesis that $f(x)$ itself is summable, let $f_N(x)$ be the function which is equal

to $f(x)$ when $|f(x)| \leqq N$, and equal to 0 when $|f(x)| > N$. Let a_k, b_k be the Fourier coefficients of $f(x)$, and a_{kN}, b_{kN} the corresponding Fourier coefficients of $f_N(x)$. It is known from the earlier work that $\lim_{k=\infty} a_{kN} = 0$, $\lim_{k=\infty} b_{kN} = 0$, for fixed N, since $f_N(x)$ is bounded. Let ε be an arbitrary positive quantity, and let N be taken so large that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} |f(x) - f_N(x)|\, dx < \frac{1}{2}\varepsilon.$$

The hypothesis of summability implies that such a choice of N is possible. Then

$$\frac{1}{\pi}\int_{-\pi}^{\pi} |[f(x) - f_N(x)] \cos kx|\, dx < \frac{1}{2}\varepsilon,$$

$$\frac{1}{\pi}\int_{-\pi}^{\pi} |[f(x) - f_N(x)] \sin kx|\, dx < \frac{1}{2}\varepsilon,$$

for all values of k, and hence

$$|a_k - a_{kN}| < \frac{1}{2}\varepsilon, \qquad |b_k - b_{kN}| \leqq \frac{1}{2}\varepsilon.$$

But for the particular value of N in question there is a k_0 such that $|a_{kN}| \leqq \frac{1}{2}\varepsilon$, $|b_{kN}| \leqq \frac{1}{2}\varepsilon$, for $k \geqq k_0$, and for values of k subject to the last condition it follows that $|a_k| < \varepsilon$, $|b_k| \leqq \varepsilon$. *If $f(x)$ is a summable function of period 2π, and if a_k, b_k are its Fourier coefficients,*

$$\lim_{k=\infty} a_k = 0, \qquad \lim_{k=\infty} b_k = 0.$$

Now let $f(x)$, still summable and of period 2π, be supposed to vanish identically for $x_0 - \eta \leqq x \leqq x_0 + \eta$, $0 < \eta < \pi$. As was shown in the first chapter,

$$S_n(x_0) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t) \frac{\sin(n+\frac{1}{2})(t-x_0)}{2\sin\frac{1}{2}(t-x_0)}\, dt.$$

Let $t - x_0 = u$; inasmuch as the integral of a periodic function over a period is the same, wherever the initial point is taken, the limits $-\pi$, π may be retained after the substitution:

$$S_n(x_0) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x_0+u)\,\frac{\sin(n+\frac{1}{2})u}{2\sin\frac{1}{2}u}\,du,$$

or, after expansion of $\sin(nu+\frac{1}{2}u)$ by the addition theorem,

$$S_n(x_0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x_0+u)\cot\frac{1}{2}u\sin nu\,du$$
$$+\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x_0+u)\cos nu\,du.$$

The factor $\cot\frac{1}{2}u$ is bounded and continuous over the range of integration, outside the interval where $f(x_0+u)$ vanishes identically. Hence $f(x_0+u)\cot\frac{1}{2}u$, like $f(x_0+u)$ itself, is a summable function (or a bounded function integrable in the sense of Riemann, if this hypothesis was originally imposed on $f(x)$), and the reasoning of the preceding paragraph is directly applicable to show that both integrals in the last expression approach zero as n becomes infinite; that is,

$$\lim_{n=\infty} S_n(x_0) = 0.$$

More generally, if $f(x)$ and $\varphi(x)$ are two summable functions (or two functions satisfying the alternative hypothesis), if $S_n(x)$ and $s_n(x)$ are their respective Fourier sums of the nth order, and if $f(x)$ and $\varphi(x)$ are identically equal for $x_0-\eta \leq x < x_0+\eta$, then

$$\lim_{n=\infty}[S_n(x_0)-s_n(x_0)] = 0;$$

if the Fourier series for $\varphi(x)$ *converges for* $x = x_0$ *to the value* $\varphi(x_0) = f(x_0)$, *the Fourier series for* $f(x)$ *does the same.* This is commonly expressed by saying that *the convergence of the Fourier series for a given function at a specified point depends only on the behavior of the function in the neighborhood of the point;* it is tacitly understood that only functions of some specified class are considered, as, in the present connection, summable functions, or functions which are bounded and integrable.

It is possible then to pass immediately from Corollary IIa of Theorem IX in the first chapter to the following statement:

THEOREM I. *If $f(x)$ is a summable function of period 2π* (or, less generally, a function of period 2π which is bounded and integrable in the sense of Riemann), *and if there is an interval $(x_0 - \eta,\ x_0 + \eta)$, $0 < \eta < \pi$, throughout which $f(x)$ is continuous, with a modulus of continuity $\omega(\delta)$ such that* $\lim_{\delta=0} \omega(\delta) \log \delta = 0$, *then the Fourier series for $f(x)$ converges for $x = x_0$ to the value $f(x_0)$.*

For a periodic function $\varphi(x)$ can be defined as equal to $f(x)$ in $(x_0 - \eta,\ x_0 + \eta)$, and linear, say, from $x_0 + \eta$ to $x_0 - \eta + 2\pi$, and this $\varphi(x)$ will have a modulus of continuity satisfying the requirements of the Corollary cited.

Attention will next be directed to the simplest case of convergence at a point of discontinuity. Let $f(x)$ now be a function which has a "finite jump", or discontinuity of the first kind, at the point $x = x_0$, approaching limits which may be denoted by $f(x_0 -)$ and $f(x_0 +)$ as x approaches x_0 from the left and from the right respectively. Let it be supposed that the values of $f(x)$ in the interval $x_0 - \eta \leq x < x_0$, together with the value $f(x_0 -)$, form a continuous function with modulus of continuity $\omega_1(\delta)$, and that $f(x)$ is likewise continuous, with the other limiting value at x_0, in the interval from x_0 to $x_0 + \eta$, the modulus of continuity this time being $\omega_2(\delta)$. For each value of δ, let $\omega(\delta)$ represent the larger of the numbers ω_1, ω_2. It will be said briefly that *$f(x)$ is continuous for $x_0 - \eta \leq x \leq x_0 + \eta$, with modulus of continuity $\omega(\delta)$, except for a finite jump at the point x_0.* It is understood always that $f(x)$ is summable over a period. If $\lim_{\delta=0} \omega(\delta) \log \delta = 0$, Theorem I establishes the convergence of the Fourier series for $f(x)$ at all interior points of the interval $(x_0 - \eta,\ x_0 + \eta)$, other than x_0, and it remains to consider the question of convergence at the point x_0 itself.

For this purpose, it may be assumed without loss of generality that $x_0 = 0$. For the terms of the series which involve $\cos nx$ and $\sin nx$, taken together, are identical with the corresponding terms in the development of $f(x)$ as a function

of $x - x_0$. In formulas, let $x - x_0 = y$, $f(x) = \varphi(y)$, $t - x_0 = u$. Then $u - y = t - x$, and

$$\begin{aligned}
&\cos ny \int_{-\pi}^{\pi} \varphi(u) \cos nu\, du + \sin ny \int_{-\pi}^{\pi} \varphi(u) \sin nu\, du \\
&= \int_{-\pi}^{\pi} \varphi(u) \cos n(u - y)\, du = \int_{-\pi}^{\pi} f(t) \cos n(t - x)\, dt \\
&= \cos nx \int_{-\pi}^{\pi} f(t) \cos nt\, dt + \sin nx \int_{-\pi}^{\pi} f(t) \sin nt\, dt.
\end{aligned}$$

The assumption that $x_0 = 0$ is therefore equivalent to a change of variable which does not affect the conditions of the problem.

With this understanding, let the symbols $f(x_0+)$ and $f(x_0-)$ be replaced by $f(0+)$ and $f(0-)$, let

$$f_1(x) = f(x) \quad (0 < x \leq \pi), \qquad f_1(x) = f_1(-x) \quad (-\pi \leq x < 0),$$
$$f_1(0) = f(0+),$$

and let $f_2(x)$ similarly be an even function identical with $f(x)$ for $-\pi \leqq x < 0$, and taking on the value $f(0-)$ for $x = 0$. Let $S_{n1}(x)$ and $S_{n2}(x)$ be the partial Fourier sums for $f_1(x)$ and $f_2(x)$ respectively. The function $f_1(x)$ is continuous throughout the interval $-\eta \leqq x \leqq \eta$. If x_1 and x_2 are two numbers belonging to this interval, and if $|x_2 - x_1| \leq \delta$,

$$|f_1(x_2) - f_1(x_1)| \leqq \omega(\delta)$$

if x_1 and x_2 have the same sign, or if one of them is zero, and

$$|f_1(x_2) - f_1(x_1)| \leqq |f_1(x_2) - f_1(0)| + |f_1(0) - f_1(x_1)| \leq 2\,\omega(\delta)$$

if x_1 and x_2 have opposite signs. Similarly, $f_2(x)$ has a modulus of continuity $\leq 2\,\omega(\delta)$ for $-\eta \leqq x \leqq \eta$. If it is supposed that $\lim_{\delta=0} \omega(\delta) \log \delta = 0$,

$$\lim_{n=\infty} S_{n1}(0) = f_1(0) = f(0+), \quad \lim_{n=\infty} S_{n2}(0) = f_2(0) = f(0-),$$

by Theorem I. But

$$S_{n1}(0) = \frac{1}{\pi}\int_{-\pi}^{\pi} f_1(t)\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt = \frac{2}{\pi}\int_0^{\pi} f(t)\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt,$$

$$S_{n2}(0) = \frac{1}{\pi}\int_{-\pi}^{\pi} f_2(t)\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt = \frac{2}{\pi}\int_{-\pi}^{0} f(t)\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt,$$

$$S_n(0) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt = \frac{1}{2}\,[S_{n1}(0)+S_{n2}(0)],$$

and therefore

$$\lim_{n\to\infty} S_n(0) = \frac{1}{2}[f(0+)+f(0-)].$$

The conclusion may be appended to the theorem as

Corollary I. *If $f(x)$ satisfies the conditions of Theorem I, except for a finite jump at the point x_0, the series converges at x_0 to the value $\frac{1}{2}[f(x_0+)+f(x_0-)]$.*

So far, nothing has been said about uniformity of convergence. To lead up to a discussion of this topic, the following lemma will be established:

Lemma I. *If $f(x)$ is an arbitrary summable function, the integrals*

$$A_k(y) = \int_{-\pi}^{y} f(t)\cos kt\,dt, \qquad B_k(y) = \int_{-\pi}^{y} f(t)\sin kt\,dt,$$

approach zero, as k becomes infinite, uniformly for $-\pi \leq y \leq \pi$.

If $f(x)$ is bounded and integrable in the sense of Riemann, the proof holds without change of form when the integrals are thought of as Riemann integrals.

In the first place, it is clear that

$$\lim_{k\to\infty} A_k(y) = \lim_{k\to\infty} B_k(y) = 0$$

for any *fixed* value of y in the interval. For $A_k(y)/\pi$ and $B_k(y)/\pi$ are the Fourier coefficients of a function $\psi(x)$ which is equal to $f(x)$ for $-\pi < x \leq y$, and equal to 0 for $y < x \leq \pi$. *Let ϵ be any positive quantity.* Let

$$A(y) = \int_{-\pi}^{y} |f(t)|\,dt.$$

This function is continuous for $-\pi \leq y \leq \pi$, and so uniformly continuous. Let δ be a positive number such that

$$|A(y_2) - A(y_1)| < \frac{1}{2}\varepsilon$$

if y_1 and y_2 are any two points in $(-\pi, \pi)$ for which $|y_2 - y_1| < \delta$. Let N be an integer, for definiteness the smallest integer, such that $2\pi/N < \delta$, and let $z_j = -\pi + (2j\pi/N)$, $j = 0, 1, \cdots, N$. By the remark made above as to the behavior of $A_k(y)$ and $B_k(y)$ for fixed y, there is for each value of j a number k_j such that $|A_k(z_j)| < \frac{1}{2}\varepsilon$, $|B_k(z_j)| < \frac{1}{2}\varepsilon$, for $k \geqq k_j$. *Let k' be the largest of the numbers $k_0, \cdots, k_N$.* If y has any value in the interval $(-\pi, \pi)$, there is a z_j such that $0 \leqq y - z_j < \delta$. Then, for any value of k,

$$|A_k(y) - A_k(z_j)| = \left| \int_{z_j}^{y} f(t) \cos k t \, d t \right| \leqq \int_{z_j}^{y} |f(t)| \, d t$$

$$= A(y) - A(z_j) < \frac{1}{2}\varepsilon,$$

$$|B_k(y) - B_k(z_j)| < \frac{1}{2}\varepsilon.$$

If $k \geqq k'$,

$$|A_k(y)| \leqq |A_k(z_j)| + |A_k(y) - A_k(z_j)| < \varepsilon, \quad |B_k(y)| < \varepsilon.$$

This is equivalent to the conclusion of the lemma.

The existence of the period 2π for $f(x)$ being understood throughout, let the definition of $A_k(y)$ and $B_k(y)$ be extended to all real values of y. For any y_1 and y_2,

$$\int_{y_1}^{y_2} f(t) \cos kt \, dt = A_k(y_2) - A_k(y_1).$$

Here and subsequently, each relation written down for A_k has a counterpart involving B_k. Let ε_k be defined for each k as the larger of the maximum values attained by $|A_k(y)|$, $|B_k(y)|$ in $(-\pi, \pi)$. In particular, $|A_k(\pi)| \leqq \varepsilon_k$, $|B_k(\pi)| \leqq \varepsilon_k$. The lemma just proved is equivalent to the statement that $\lim_{k=\infty} \varepsilon_k = 0$. Because of the periodicity of $f(t) \cos kt$ and $f(t) \sin kt$, the integrals defining $A_k(y)$ and $B_k(y)$ are not altered in value if both limits of integration are increased

or diminished by the same integral multiple of 2π. For example, if y is in the interval $(\pi, 3\pi)$,

$$|A_k(y) - A_k(\pi)| \leqq \varepsilon_k,$$

whence it follows that $|A_k(y)| \leqq 2\varepsilon_k$ in this interval. If y_1 and y_2 are anywhere in $(-\pi, 3\pi)$, the corresponding values of A_k will differ by not more than $4\varepsilon_k$. Finally, if y_1 and y_2 are any two numbers subject to the condition that $|y_2 - y_1| \leqq 2\pi$, there will be an integer j (positive, negative, or zero) such that $y_1 + 2j\pi$ and $y_2 + 2j\pi$ belong to the interval $(-\pi, 3\pi)$, and

$$|A_k(y_2) - A_k(y_1)| \leqq 4\varepsilon_k.$$

A similar relation holds for B_k.

The purpose of these details is to bring out a fact which will be used presently. For any x, and for any y in $(-\pi, \pi)$,

$$\left|\int_{-\pi+x}^{y+x} f(t)\sin n(t-x)\,dt\right|$$
$$= |\cos nx[B_n(x+y) - B_n(x-\pi)] - \sin nx[A_n(x+y) - A_n(x-\pi)]|$$
$$\leqq 8\varepsilon_n,$$

or, by the substitution $t - x = u$,

$$\left|\int_{-\pi}^{y} f(x+u)\sin nu\,du\right| \leqq 8\varepsilon_n$$

for arbitrary x and y, subject to the condition that $-\pi \leqq y \leqq \pi$. (It is clear that the interval from $-\pi$ to y may be replaced by any other interval of length $\leqq 2\pi$.)

The next stage of the discussion may be summarized in another lemma:

LEMMA II. *If $f(x)$ is a summable function which vanishes identically for $\alpha - \eta \leqq x \leqq \beta + \eta$, with $\eta > 0$, its Fourier series converges uniformly to the value 0 for $\alpha \leqq x \leqq \beta$.*

The proof will be expressed in terms of the Lebesgue theory of integration, and, unlike those that have gone before, would have to be appreciably modified in form if reference to that theory were to be eliminated. A more elementary method of proof appropriate to the case of functions bounded and integrable in the sense of Riemann will be indicated later.

Let x have a value belonging to the interval $\alpha \leqq x \leqq \beta$. By the usual formula,

$$S_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t) \frac{\sin(n+\frac{1}{2})(t-x)}{2\sin\frac{1}{2}(t-x)}\,dt$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\cot\frac{1}{2}(t-x)\sin n(t-x)\,dt$$

$$+ \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\cos n(t-x)\,dt.$$

The second of the two terms making up the last expression is equal to $\frac{1}{2}(a_n \cos nx + b_n \sin nx)$, and approaches zero uniformly for all values of x, as an immediate consequence of the fact that a_n and b_n approach zero.

In the first term, let $t - x = u$. Then the integral is equal to

$$J_n(x) = \int_{-\pi}^{\pi} f(x+u)\cot\frac{1}{2}u \sin nu\,du.$$

Let $C(u)$ be a function of period 2π which is equal to $\cot\frac{1}{2}u$ for $-\pi \leqq u \leqq -\eta$ and for $\eta \leqq u \leqq \pi$, equal to 0 for $-\frac{1}{2}\eta \leqq u \leqq \frac{1}{2}\eta$, and so defined for $-\eta \leqq u \leqq -\frac{1}{2}\eta$ and for $\frac{1}{2}\eta \leqq u \leqq \eta$ as to have a continuous derivative everywhere; it is scarcely necessary to write down an explicit formula for the construction. As $f(x+u)$ is identically zero for $-\eta \leqq u \leqq \eta$, when x is in (α, β), the integral $J_n(x)$ is the same as

$$\int_{-\pi}^{\pi} f(x+u)\,C(u)\sin nu\,du.$$

Let

$$\int_{-\pi}^{y} f(x+u)\sin nu\,du = F(y);$$

the values of n and x are for the moment to be regarded as fixed. The function $F(y)$ has almost everywhere a derivative equal to $f(x+y)\sin ny$. Consequently, as $C(u)$ is differentiable everywhere, the function $F(u)\,C(u)$ has almost everywhere a derivative equal to $F'(u)\,C(u) + F(u)\,C'(u)$. Furthermore, $F(u)$ is absolutely continuous, and $C(u)$, having

a bounded derivative, is absolutely continuous also. Hence $F(u)\,C(u)$ is absolutely continuous, and its increment over an interval is equal to the integral of its derivative over the interval. That is, in particular,

$$F(\pi)\,C(\pi) - F(-\pi)\,C(-\pi)$$
$$= \int_{-\pi}^{\pi} F'(u)\,C(u)\,du + \int_{-\pi}^{\pi} F(u)\,C'(u)\,du,$$

$$J_n(x) = \int_{-\pi}^{\pi} F'(u)\,C(u)\,du$$
$$= F(\pi)\,C(\pi) - F(-\pi)\,C(-\pi) - \int_{-\pi}^{\pi} F(u)\,C'(u)\,du.$$

But $C(\pi) = C(-\pi) = \cot(\pm\frac{1}{2}\pi) = 0$. By an earlier paragraph, $|F(u)| \leq 8\epsilon_n$ for $-\pi \leqq u \leqq \pi$, where ϵ_n is independent of x; if μ is the maximum of the continuous function $|C'(u)|$, which is independent of x and n,

$$|J_n(x)| = \left|\int_{-\pi}^{\pi} F(u)\,C'(u)\,du\right| \leq 16\pi\mu\epsilon_n.$$

So $J_n(x)$ approaches zero uniformly for the values of x in question, $S_n(x)$ does likewise, and the lemma is proved.

It follows at once that if $f(x)$ and $\varphi(x)$ are two summable functions, identically equal for $\alpha - \eta \leqq x \leqq \beta + \eta$, and if the Fourier series for $\varphi(x)$ converges uniformly to the value $\varphi(x) = f(x)$ for $\alpha \leqq x \leqq \beta$, the series for $f(x)$ does the same. In particular, Theorem I may be further supplemented by

COROLLARY II. *If $f(x)$ is a summable function of period 2π, and if there is an interval $\alpha - \eta \leqq x \leqq \beta + \eta$, $\eta > 0$, throughout which $f(x)$ is continuous, with a modulus of continuity $\omega(\delta)$ such that* $\lim_{\delta=0} \omega(\delta) \log \delta = 0$, *then the Fourier series for $f(x)$ converges uniformly to the value $f(x)$ for $\alpha \leqq x \leqq \beta$.*

2. Convergence of Fourier series under hypothesis of limited variation

As the next pages will be concerned largely with functions of limited variation, the insertion of a proof of the theorem about the convergence of the Fourier series for such a function is not out of place.

Let $f(x)$ be a function of period 2π, having limited variation over any finite interval. Since a shift of the origin from which x is measured does not change either the character of the function or the terms of its Fourier series, when sine and cosine terms of like order are taken together, it is sufficient to consider convergence for $x = 0$.

Since $f(x)$ is of limited variation, its discontinuities, if any, are finite jumps. Let $\psi(x)$ be a function of period 2π which is linear for $0 < x < 2\pi$, and further defined by the conditions $\psi(0+) = f(0+)$, $\psi(0-) = \psi(2\pi-) = f(0-)$, $\psi(0) = \frac{1}{2}[f(0+)+f(0-)]$. If $f(x)$ is continuous for $x = 0$, $\psi(x)$ is merely a constant. If $f(x)$ is discontinuous for $x = 0$, $\psi(x)$ is equal to the constant $\frac{1}{2}[f(0+)+f(0-)]$ plus an odd function, and its Fourier series consists of this constant plus a series of sine terms, all of which vanish at the origin. So the partial sum of this Fourier series not merely approaches $\psi(0)$ (in accordance with Corollary I above) but is always exactly equal to $\psi(0)$. As $f(x) = \psi(x) + [f(x) - \psi(x)]$, the problem of convergence for $f(x)$ reduces immediately to the corresponding problem for the difference $f(x) - \psi(x)$, which is continuous at the origin, if $f(x)$ is defined for $x = 0$ as equal to the mean of its limiting values. There is no loss of generality therefore in assuming at the outset that $f(x)$ is continuous at the origin and vanishes there. This assumption will be made henceforth.

By the hypothesis of limited variation, $f(x)$ can be expressed for $-\pi \leqq x \leqq \pi$ in the form $f(x) = \varphi_1(x) - \varphi_2(x)$, where φ_1 and φ_2 are bounded and monotone increasing throughout the interval, are continuous wherever $f(x)$ is continuous, for $x = 0$ in particular, and vanish for $x = 0$. For the study of the expression

$$S_n(0) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt,$$

let

$$j_{11} = \int_0^{\pi} \varphi_1(t)\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt,$$

let j_{12} be the integral similarly formed with φ_2 in place of φ_1, and let j_{21} and j_{22} be the corresponding integrals from $-\pi$ to 0. Then $S_n(0) = (1/\pi)(j_{11}-j_{12}+j_{21}-j_{22})$.

If ξ_1 and ξ_2 are any two numbers belonging to the interval $(0, \pi)$, and if n is any positive integer,

$$\left|\int_{\xi_1}^{\xi_2} \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt\right| \leqq 3\pi.$$

For the integrand is alternately positive and negative over intervals of length $\pi/(n+\frac{1}{2})$, the arches of its graph diminishing steadily in height from left to right. The integral over an interval corresponding to any number of whole arches is therefore a sum of diminishing terms of alternate signs, and its magnitude does not exceed that of the largest term, namely the area of the first arch involved. The integral from any ξ_1 to any ξ_2 is made up at worst of such a sum, plus the integrals corresponding to parts of two other arches. But the integrand, being equal to $\frac{1}{2}+\cos t+\cdots+\cos nt$, never exceeds $n+\frac{1}{2}$ in absolute value, and the area of any one arch, with base $\pi/(n+\frac{1}{2})$, can not exceed π. So the magnitude of the integral from ξ_1 to ξ_2 can not exceed 3π. (It is almost as easy to see that the value is actually less than 2π, and a still lower bound could be obtained, but the inequality as written is sufficient for the purpose in hand; the essential thing is that the right-hand member is independent of ξ_1, ξ_2, and n.)

If $0<\delta \leqq \xi_1 \leqq \xi_2 \leqq \pi$, the number δ being regarded as fixed,

$$\left|\int_{\xi_1}^{\xi_2} \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt\right| \leqq \frac{3\pi c_\delta}{n+\frac{1}{2}}, \quad c_\delta = \frac{1}{2\sin\frac{1}{2}\delta}.$$

For the absolute value of the integrand never exceeds c_δ, and the magnitude of the integral can not exceed the sum of the areas of three arches, each of base $\pi/(n+\frac{1}{2})$ and of height not greater than c_δ.

Let ϵ be any positive quantity, and let δ be chosen (by virtue of the continuity and vanishing of φ_1 at the origin)

so that $\varphi_1(\delta) \leqq \varepsilon/(6\pi)$. The integral defining j_{11} may be taken as the sum of the integrals from 0 to δ and from δ to π. The second law of the mean may be applied in each of these integrals. For that from 0 to δ it gives

$$\int_0^\delta \varphi_1(t) \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\, dt = \varphi_1(\delta) \int_\xi^\delta \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\, dt,$$

where ξ has some value in the interval $(0, \delta)$; by the second paragraph preceding, the absolute value of the integral on the right does not exceed 3π, and consequently that of the integral on the left does not exceed $\frac{1}{2}\varepsilon$, for any value of n. For the rest of j_{11},

$$\int_\delta^\pi \varphi_1(t) \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\, dt = \varphi_1(\delta) \int_\delta^\xi \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\, dt$$
$$+ \varphi_1(\pi) \int_\xi^\pi \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\, dt,$$

the new ξ being in the interval (δ, π); here $\varphi_1(\delta) \leqq \varphi_1(\pi)$, and the absolute value of each integral on the right, by the preceding paragraph, is less than or equal to $3\pi c_\delta/(n+\frac{1}{2})$. So

$$\left|\int_\delta^\pi \varphi_1(t) \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\, dt\right| < \frac{6\pi c_\delta \varphi_1(\pi)}{n+\frac{1}{2}},$$

which is less than $\frac{1}{2}\varepsilon$ as soon as n is sufficiently large. This means that j_{11} approaches zero as n becomes infinite. By similar treatment of j_{12}, j_{21}, and j_{22}, it is recognized that $S_n(0)$ approaches the value $0 = f(0)$.

For functions having the original degree of generality, the result may be stated in the following form:

THEOREM II. *If $f(x)$ is a function of period 2π having limited variation over a period, its Fourier series converges to the value $f(x)$ at every point where $f(x)$ is continuous, and to the mean of the values approached from the right and from the left at every point where $f(x)$ is discontinuous.*

If $f(x)$ is discontinuous at the origin, the details of the above calculation apply properly not to $f(x)$ itself, but to

the difference $f(x) - \psi(x)$, the retention of the symbol $f(x)$ having been equivalent to a change of notation. Let it be supposed now that $f(x)$ is continuous throughout. Then the complication just mentioned does not arise (except for the subtraction of a constant to make $f(0) = 0$, which is of no consequence). The functions $\varphi_1(x)$ and $\varphi_2(x)$ may be taken as the positive and negative variations of $f(x)$ itself (in the sense that $\varphi_1(x)$ is the positive variation from 0 to x when x is positive and minus the positive variation from x to 0 when x is negative, φ_2 being similarly defined). The use of the origin as representative of an arbitrary point x_0 in the convergence proof amounts to another change of notation, in connection with which $\varphi_1(\delta)$, in the course of the reasoning, takes the place of $\varphi_1(x_0 + \delta) - \varphi_1(x_0)$. Under the present hypotheses, φ_1 and φ_2 are everywhere continuous, and so uniformly continuous. Hence the choice of δ, if the proof is written out in terms of the general notation, can be made independently of x_0. For any x_0, furthermore, $\varphi_1(x_0 + \pi) - \varphi_1(x_0)$ and $\varphi_2(x_0 + \pi) - \varphi_2(x_0)$ can not exceed the total variation of $f(x)$ over a period, a quantity likewise independent of x_0. These are the essential points needed to justify the supplementary assertion:

COROLLARY I. *If $f(x)$ satisfies the hypotheses of Theorem* II, *and is furthermore continuous everywhere, the Fourier series converges to $f(x)$ uniformly for all values of x.*

If it is assumed merely that $f(x)$ is of limited variation for $x_0 - \eta \leqq x < x_0 + \eta$, and summable over a period, the function $\varphi(x)$ which is equal to $f(x)$ for $x_0 - \eta \leqq x < x_0 + \eta$, and identically zero over the rest of a period, is of limited variation over the entire period, and its Fourier series converges at x_0 to the value $\frac{1}{2}[\varphi(x_0+) + \varphi(x_0-)] = \frac{1}{2}[f(x_0+) + f(x_0-)]$, from which it follows that the Fourier series for $f(x)$ does the same. If $f(x)$ is continuous and of limited variation for $\alpha - \eta \leqq x \leqq \beta + \eta$, and summable over a period, the function $\varphi(x)$ of period 2π which is equal to $f(x)$ for $\alpha - \eta \leqq x \leqq \beta + \eta$, and linear for $\beta + \eta \leqq x \leqq \alpha - \eta + 2\pi$, is of limited variation and continuous over the entire period,

and its Fourier series converges uniformly with $\varphi(x)$ for its sum; the Fourier series for $f(x)-\varphi(x)$ converges uniformly to zero for $\alpha \leqq x \leqq \beta$, by Lemma II; and consequently the series for $f(x)$ converges uniformly to $f(x)$ for $\alpha \leqq x \leqq \beta$. These conclusions may be expressed as

COROLLARY II. *If $f(x)$ is a function of period 2π, summable over a period, and of limited variation for $x_0-\eta \leqq x \leqq x_0+\eta$, its Fourier series converges for $x = x_0$ to the value $\frac{1}{2}[f(x_0+)+f(x_0-)]$; if $f(x)$ is of period 2π, summable over a period, and continuous and of limited variation for $\alpha-\eta \leqq x \leqq \beta+\eta$, its Fourier series converges uniformly to the value $f(x)$ for $\alpha \leqq x \leqq \beta$.*

It may be noted in passing that the second law of the mean leads to a proof of Lemma II for functions that are bounded and integrable in the sense of Riemann, without the use of Lebesgue integration. In the expression which was denoted by $J_n(x)$, in the proof as given previously, let the integrals from $-\pi$ to $-\eta$ and from η to π be considered separately; the integrand is identically zero for $-\eta \leqq u \leqq \eta$. Since $\cot \frac{1}{2}u$ is monotone from η to π, and $\cot \frac{1}{2}\pi = 0$,

$$\int_\eta^\pi f(x+u)\cot\frac{1}{2}u \sin nu\,du = \left(\cot\frac{1}{2}\eta\right)\int_\eta^\xi f(x+u)\sin nu\,du,$$

where ξ is a number of the interval (η, π). The magnitude of the last integral does not exceed the quantity $8\epsilon_n$, which is independent of x and approaches zero as n becomes infinite. Similarly it may be shown that the integral from $-\pi$ to $-\eta$ approaches zero uniformly, and the conclusion of the lemma follows at once.

3. Degree of convergence of Fourier series under hypotheses involving limited variation

Attention will now be directed once more to questions of degree of convergence. The next theorem is a rather simple one:

THEOREM III. *If $f(x)$ is a function of period 2π with limited variation, the total variation over a period being V, then, for $n > 0$,*

$$\left|\int_{-\pi}^{\pi} f(x)\cos nx\,dx\right| \leqq \frac{V}{n}, \qquad \left|\int_{-\pi}^{\pi} f(x)\sin nx\,dx\right| \leqq \frac{V}{n}.$$

Let $\varphi_1(x)$ and $\varphi_2(x)$ be the positive and negative variations of $f(x)$, starting from the point $-\pi$, so that $f(x) = f(-\pi) + \varphi_1(x) - \varphi_2(x)$, $\varphi_1(-\pi) = \varphi_2(-\pi) = 0$. Since $f(\pi) = f(-\pi)$, $\varphi_1(\pi) - \varphi_2(\pi) = 0$. But $\varphi_1(\pi) + \varphi_2(\pi) = V$, and therefore $\varphi_1(\pi) = \varphi_2(\pi) = \frac{1}{2}V$. By the second law of the mean,

$$\int_{-\pi}^{\pi} \varphi_1(x)\cos nx\,dx = \varphi_1(\pi)\int_{\xi}^{\pi}\cos nx\,dx$$

$$= \frac{1}{2}V\left(\frac{\sin n\pi - \sin n\xi}{n}\right) = -\frac{V\sin n\xi}{2n},$$

$$\left|\int_{-\pi}^{\pi} \varphi_1(x)\cos nx\,dx\right| \leqq \frac{V}{2n}.$$

Similarly,

$$\left|\int_{-\pi}^{\pi} \varphi_2(x)\cos nx\,dx\right| \leqq \frac{V}{2n},$$

while

$$\int_{-\pi}^{\pi} f(-\pi)\cos nx\,dx = 0.$$

So the first inequality of the theorem is obtained. A mechanical repetition of the reasoning for the integral with $\sin nx$ would lead to the expression $\cos n\xi - \cos n\pi$ in place of $\sin n\pi - \sin n\xi$, and the new expression has the maximum value 2 instead of 1, since $\cos n\pi \neq 0$. But by virtue of the periodicity the integral from $-\pi$ to π is the same as that from $-\pi + [\pi/(2n)]$ to $\pi + [\pi/(2n)]$, and if the variations are measured from the left-hand end of the latter interval the upper bound V/n is obtained once more, as stated in the theorem.

Suppose now that $f(x)$ has a first derivative with limited variation, and let V be the total variation of $f'(x)$ over a period. By integration by parts,

$$\int_{-\pi}^{\pi} f(x)\cos nx\,dx = \left[\frac{f(x)\sin nx}{n}\right]_{-\pi}^{\pi} - \frac{1}{n}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx$$

$$= -\frac{1}{n}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx,$$

whence it follows further, since Theorem III is applicable to the last integral, that

$$\left|\int_{-\pi}^{\pi} f(x)\cos nx\,dx\right| = \frac{1}{n}\left|\int_{-\pi}^{\pi} f'(x)\sin nx\,dx\right| \leq \frac{V}{n^2}.$$

Similar reasoning applies to the integral with $\sin nx$ in place of $\cos nx$. Repetition of the process of integration by parts leads to the following:

COROLLARY I. *If $f(x)$ is a function of period 2π which has a pth derivative with limited variation, $p > 0$, and if V is the total variation of $f^{(p)}(x)$ over a period, then, for $n > 0$,*

$$\left|\int_{-\pi}^{\pi} f(x)\cos nx\,dx\right| \leq \frac{V}{n^{p+1}}, \qquad \left|\int_{-\pi}^{\pi} f(x)\sin nx\,dx\right| \leq \frac{V}{n^{p+1}}.$$

The conclusion is really somewhat more general than the statement would indicate. Suppose for simplicity once more that $p = 1$. As will be seen on re-examination of the proof in the light of well known theorems on Lebesgue integrals, the essential thing is not that $f'(x)$ be uniquely defined at every point, but that $f(x)$ be expressible as the integral of a function $\varphi(x)$ of limited variation:

$$f(x) = f(a) + \int_a^x \varphi(x)\,dx,$$

for any value of a. This observation is of some interest, since the simplest functions represented by graphs with corners satisfy the modified hypothesis, but not the original one. The generalization carries over to the applications of the Corollary in Theorems IV and Vb below.

For the sake of another corollary, let $f(x)$ be identically zero for $x_0 - \eta \leq x \leq x_0 + \eta$, and of limited variation over a period. If $S_n(x_0)$ is represented once more by the formula

$$S_n(x_0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x_0+u)\cot\frac{1}{2}u\sin nu\,du$$
$$+\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x_0+u)\cos nu\,du,$$

both $f(x_0+u)$ and $f(x_0+u)\cot\frac{1}{2}u$ are of limited variation, regarded as functions of u, and consequently $|S_n(x_0)|$ does not exceed a constant multiple of $1/n$. With regard to uniformity of convergence, let $f(x)$ be identically zero for $\alpha-\eta\leqq x\leqq\beta+\eta$, the hypothesis of limited variation over a period being kept unchanged. When $\alpha\leqq x_0\leqq\beta$, the function $f(x_0+u)\cot\frac{1}{2}u$ is identical with $f(x_0+u)C_1(u)$, if $C_1(u)$ is defined as equal to 0 for $0\leqq|u|<\eta$, and equal to $\cot\frac{1}{2}u$ for $\eta\leqq|u|\leqq\pi$. In general, if $\varphi_1(u)$ and $\varphi_2(u)$ are any two functions of limited variation, the total variation of $\varphi_1(u)$ being V_1 and that of $\varphi_2(u)$ being V_2, and if M_1 and M_2 are upper bounds for $|\varphi_1(u)|$ and $|\varphi_2(u)|$ respectively, the function $\varphi_1(u)\varphi_2(u)$ is of limited variation, and its total variation does not exceed $M_1V_2+M_2V_1$. In the present instance, $\varphi_1(u)$ and $\varphi_2(u)$ being replaced by $f(x_0+u)$ and $C_1(u)$ respectively, $f(x_0+u)$ has the same total variation over a period and the same bounds as $f(x)$, for any value of x_0, and $C_1(u)$ is independent of x_0, so that the total variation of $f(x_0+u)\cot\frac{1}{2}u$, while presumably different for different values of x_0, has an upper bound independent of x_0, as long as x_0 belongs to the interval (α,β). Such an upper bound can be calculated more specifically as the product of V, the total variation of $f(x)$, by a quantity depending only on η. (The absolute value of $f(x)$ can not exceed V anywhere, since every period contains points where $f(x)$ vanishes.) It is possible therefore to state

COROLLARY II. *If $f(x)$ is a function of period 2π with limited variation, the total variation over a period being V, and if $f(x)$ vanishes identically for $\alpha-\eta\leqq x\leqq\beta+\eta$, then*

$$|S_n(x)|\leqq\frac{C_\eta V}{n}$$

for $\alpha\leqq x\leqq\beta$, where $S_n(x)$ is the partial sum of the Fourier series for $f(x)$, and C_η depends only on η.

Each of these corollaries, taken in conjunction with results obtained earlier, leads at once to a theorem on degree of convergence. If $p\geqq1$ in the hypothesis of Corollary I, it

is known that the Fourier series for $f(x)$ converges uniformly to the value $f(x)$; this follows from Theorem IX of Chapter I, and more directly from Corollary I or Corollary IIa of that theorem, since the hypothesis provides $f(x)$ with a bounded first derivative, and

$$|f(x_2)-f(x_1)| \leqq \lambda |x_2-x_1|,$$

if λ is an upper bound for $|f'(x)|$. On the other hand, Corollary I, interpreted in terms of the Fourier coefficients, states that

$$|a_n| \leqq V/(\pi n^{p+1}), \qquad |b_n| \leqq V/(\pi n^{p+1}).$$

Hence

$$\begin{aligned}|f(x)-S_n(x)| &= \left|\sum_{k=n+1}^{\infty}(a_k \cos kx + b_k \sin kx)\right| \\ &\leqq \sum_{k=n+1}^{\infty}(|a_k|+|b_k|) \\ &\leqq 2\sum_{k=n+1}^{\infty}\frac{V}{\pi k^{p+1}} \leqq \frac{2V}{\pi}\int_n^{\infty}\frac{du}{u^{p+1}} = \frac{2V}{p\pi n^p}.\end{aligned}$$

The conclusion may be formulated as

THEOREM IV. *If $f(x)$ is a function of period 2π which has a pth derivative with limited variation. $p \geqq 1$, and if V is the total variation of $f^{(p)}(x)$ over a period, then, for $n > 0$,*

$$|f(x)-S_n(x)| \leqq \frac{Q_p V}{n^p} \leqq \frac{QV}{n^p},$$

where $S_n(x)$ is the partial sum of the Fourier series for $f(x)$, Q_p is a constant depending only on p, and Q is an absolute constant: more specifically, $Q_p = 2/(p\pi)$, $Q = Q_1 = 2/\pi$.

This result is to be compared with Theorem X and the corollaries of Theorem IX in Chapter I.

Corollary II will be combined with Corollary II of Theorem IX in Chapter I, and with the theorem just obtained. In connection with the result from Chapter I, it is to be noted that if $f(x)$ is a continuous function which is not identically constant, and if $\omega(\delta)$ is its modulus of continuity,

$$\liminf_{\delta=0}\frac{\omega(\delta)}{\delta} > 0.$$

For if this were not the case, it would be possible for every $\epsilon > 0$ to find arbitrarily small positive values of δ such that $\omega(\delta) \leqq \epsilon\delta$; for any such δ, $|f(x_2) - f(x_1)| \leqq \epsilon|x_2 - x_1|$, if $x_2 - x_1$ is an integral multiple of δ, since $|f(x_1 + N\delta) - f(x_1)| \leqq N\omega(\delta) \leqq \epsilon \cdot N\delta$, if N is integral; but for any choice of x_1 and x_2, it would be possible to find values of δ satisfying the condition that $\omega(\delta) \leqq \epsilon\delta$, and so small that $x_2 - x_1$ differs arbitrarily little from an integral multiple of δ; and then it must still be true by continuity that $|f(x_2) - f(x_1)| \leqq \epsilon|x_2 - x_1|$, which in view of the arbitrariness of ϵ means that $f(x_2) - f(x_1) = 0$ for all x_1 and x_2, contrary to the hypothesis that $f(x)$ is not constant. The conclusion may be stated by saying that $\omega(\delta)/\delta$ has a positive lower bound for values of δ that are sufficiently small; as $\omega(\delta)$ itself has a positive lower bound over any interval not reaching to the origin, it appears further that $\omega(\delta)/\delta$ has a positive lower bound over any finite range for δ.

Suppose now that $f(x)$, having the period 2π, is continuous with modulus of continuity $\omega(\delta)$ for $\alpha - \eta \leqq x \leqq \beta + \eta$, but not constant over the interval, and of limited variation (but not necessarily continuous) over the rest of a period. Let $\varphi(x)$ be periodic with period 2π, equal to $f(x)$ for $\alpha - \eta \leqq x \leqq \beta + \eta$, and linear for $\beta + \eta \leqq x \leqq \alpha - \eta + 2\pi$. In the latter interval, the modulus of continuity of $\varphi(x)$ is a constant multiple of δ, which by the preceding remarks does not exceed a constant multiple of $\omega(\delta)$, say $k\omega(\delta)$, over the range within which $\omega(\delta)$ is defined, namely for $0 < \delta \leqq \beta - \alpha + 2\eta$. If δ satisfies the latter condition, any interval of length δ is made up at worst of an interval congruent (modulo 2π) to $(\beta + \eta, \alpha - \eta + 2\pi)$, together with parts of two intervals congruent to $(\alpha - \eta, \beta + \eta)$, and so $\varphi(x)$ has everywhere a modulus of continuity $\omega_1(\delta)$ which for $0 < \delta < \beta - \alpha + 2\eta$ does not exceed $(k+2)\,\omega(\delta)$. By the corollary cited from Chapter I, $\varphi(x)$ differs from the partial sum of its Fourier series by not more than $A\omega_1(2\pi/n)\log n$, and so by not more than $A(k+2)\,\omega(2\pi/n)\log n$, if n is large enough so that $2\pi/n$ comes within the specified range of values for δ.

On the other hand, $f(x) - \varphi(x)$ is of limited variation, and identically zero for $\alpha - \eta \leqq x \leqq \beta + \eta$, so that the recent Corollary II is applicable; and the quantity $1/n$ which enters into the conclusion of that corollary does not exceed a constant multiple of $\omega(2\pi/n)$. So it is possible to state

THEOREM Va. *If the function $f(x)$, of period 2π, is continuous with modulus of continuity $\omega(\delta)$ for $\alpha - \eta \leqq x \leqq \beta + \eta$, where $\omega(\delta) > 0$ for $\delta > 0$, and of limited variation (but not necessarily continuous) over the rest of a period. then*

$$|f(x) - S_n(x)| \leqq c\,\omega(2\pi/n) \log n$$

for $\alpha \leqq x \leqq \beta$, if n is large enough so that $\omega(2\pi/n)$ has a meaning, $S_n(x)$ being the partial sum of the Fourier series for $f(x)$, and c a constant depending neither on x nor on n.

The combination of Corollary II with Theorem IV (or rather with a part of that theorem) proceeds with a little more facility. Let $f(x)$ be a function of period 2π having a first derivative with limited variation for $\alpha - \eta \leqq x \leqq \beta + \eta$, while $f(x)$ itself is of limited variation over a period. Let $\varphi(x)$ this time be of period 2π, equal to $f(x)$ for $\alpha - \eta \leqq x \leqq \beta + \eta$, and defined for $\beta + \eta \leqq x \leqq \alpha - \eta + 2\pi$ as a polynomial of the third degree so that $\varphi(x)$ and its derivative have determinate values at both ends of the interval. Then $\varphi(x)$ has a first derivative everywhere, which is of limited variation over a period, and $f(x) - \varphi(x)$ is of limited variation over a period and identically zero for $\alpha - \eta \leqq x \leqq \beta + \eta$. It remains to apply Theorem IV to $\varphi(x)$, and Corollary II to $f(x) - \varphi(x)$, and the following theorem is obtained:

THEOREM Vb. *If the function $f(x)$, of period 2π, is of limited variation over a period, and has a first derivative of limited variation for $\alpha - \eta \leqq x \leqq \beta + \eta$, then*

$$|f(x) - S_n(x)| \leqq \frac{c}{n}$$

for $\alpha \leqq x \leqq \beta$, where $S_n(x)$ is the partial sum of the Fourier series for $f(x)$, and c is a constant depending neither on x nor on n.

Space will not be taken here for the working out of more elaborate combinations of similar character. A theorem related to Theorem III, however, will be obtained for the sake of a subsequent application.

It was shown early in the chapter that the Fourier coefficients of an arbitrary summable function of period 2π approach zero as a limit. Let $f(x)$ now be a function of period 2π which is absolutely continuous, and let a_k, b_k be its Fourier coefficients. Its derivative exists almost everywhere, and is summable. The product $f(x)\sin kx$ is likewise absolutely continuous, and so has a derivative almost everywhere, and its change of value over an interval is equal to the integral of its derivative. Specifically, it has a derivative equal to $f'(x)\sin kx + kf(x)\cos kx$ at every point where $f'(x)$ exists. Consequently

$$\begin{aligned} 0 &= f(\pi)\sin k\pi - f(-\pi)\sin(-k\pi) \\ &= \int_{-\pi}^{\pi} f'(x)\sin kx\,dx + k\int_{-\pi}^{\pi} f(x)\cos kx\,dx, \end{aligned}$$

$$k a_k = -\frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\sin kx\,dx.$$

But by the theorem cited at the beginning of the paragraph, the last integral approaches zero as k becomes infinite, since $f'(x)$ is summable. Similarly, $k b_k$ approaches zero.

It was seen at the beginning of the chapter that

$$\begin{aligned} &\frac{1}{\pi}\int_{-\pi}^{\pi}[f(x) - S_n(x)]^2\,dx \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x)]^2\,dx - \left[\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2)\right]. \end{aligned}$$

But under the hypothesis of the moment $f(x)$, being absolutely continuous, is *a fortiori* of limited variation, and hence $S_n(x)$ converges uniformly to $f(x)$. So the left-hand member approaches zero, and the right-hand member must do the same:

$$\frac{1}{\pi}\int_{-\pi}^{\pi}[f(x)]^2\,dx = \frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2).$$

(It is still not necessary for the purpose in hand to establish the well-known fact that the last relation holds under much more general hypotheses. Use of the theorem on the convergence of the Fourier series for a function of limited variation can be avoided by reference to the least-square property of the sums $S_n(x)$, discussed in the next chapter, together with the fact that by Weierstrass's theorem trigonometric sums $T_n(x)$ can be found so as to make $f(x) - T_n(x)$ approach zero uniformly.) From the last two equations, taken together, it appears that

$$\frac{1}{\pi}\int_{-\pi}^{\pi}[f(x) - S_n(x)]^2\,dx = \sum_{k=n+1}^{\infty}(a_k^2 + b_k^2).$$

Let ε be an arbitrary positive quantity. It is possible to choose n_0 so that $|k\,a_k| \leqq (\frac{1}{2}\varepsilon)^{1/2}$, $|k\,b_k| \leqq (\frac{1}{2}\varepsilon)^{1/2}$, for $k \geqq n_0$. Then, if $n \geqq n_0$,

$$\sum_{k=n+1}^{\infty}(a_k^2 + b_k^2) \leqq \varepsilon \sum_{k=n+1}^{\infty}\frac{1}{k^2} \leqq \varepsilon\int_n^{\infty}\frac{du}{u^2} = \frac{\varepsilon}{n},$$

which means that

$$\lim_{n=\infty} n \sum_{k=n+1}^{\infty}(a_k^2 + b_k^2) = 0.$$

The results of the last two paragraphs may be summarized in

THEOREM VI. *If $f(x)$ is an absolutely continuous function of period 2π, and if a_n, b_n are its Fourier coefficients,*

$$\lim_{n=\infty} n\,a_n = 0, \qquad \lim_{n=\infty} n\,b_n = 0.$$

If $S_n(x)$ is the partial sum of the Fourier series for $f(x)$, and if γ_n is defined by the equation

$$\gamma_n = \int_{-\pi}^{\pi}[f(x) - S_n(x)]^2\,dx,$$

then

$$\lim_{n=\infty} n\,\gamma_n = 0.$$

The first part of this theorem supplements Theorem III: the products $n\,a_n$, $n\,b_n$ are bounded for any function of limited

variation, and approach zero if the hypothesis of limited variation is replaced by the more stringent requirement of absolute continuity. The last part of the theorem will be used in the next chapter.

4. Convergence of the first arithmetic mean

The next undertaking will be to outline a theory of the convergence and degree of convergence of the approximating functions with which the name of Fejér is associated, the first arithmetic means of the Fourier series for a given function. The mean in question is defined by the identity

$$\sigma_n(x) = \frac{1}{n}[S_0(x) + S_1(x) + \cdots + S_{n-1}(x)],$$

where $S_k(x)$, as usual, denotes the partial sum of the Fourier series through terms of the kth order.

It will be recalled that the fundamental integral expression for $S_k(x)$ involves a factor $\sin(k+\frac{1}{2})(t-x)$, and otherwise does not change with k. The product of the expression

$$\sin\frac{1}{2}v + \sin\frac{3}{2}v + \cdots + \sin\left(n-\frac{1}{2}\right)v$$

by $2\sin\frac{1}{2}v$ can be rearranged in the form

$$[1-\cos v] + [\cos v - \cos 2v] + \cdots + [\cos(n-1)v - \cos nv]$$
$$= 1 - \cos nv = 2\sin^2(nv/2).$$

Application of this identity, with v replaced by $t-x$, gives

$$\sigma_n(x) = \frac{1}{n\pi}\int_{-\pi}^{\pi} f(t)\frac{\sin^2\frac{1}{2}n(t-x)}{2\sin^2\frac{1}{2}(t-x)}\,dt,$$

or in terms of a new variable $u = \frac{1}{2}(t-x)$:

$$\sigma_n(x) = \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2} f(x+2u)\frac{\sin^2 nu}{\sin^2 u}\,du.$$

If $(\sin^2 nu)/(\sin^2 u)$ is denoted for brevity by $\Phi_n(u)$, the expression becomes

$$\sigma_n(x) = \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2} f(x+2u)\,\Phi_n(u)\,du.$$

The fact that the trigonometric factor $\Phi_n(u)$ in the integrand, unlike the corresponding factor in the expression for $S_n(x)$, is never negative, has important consequences, and materially simplifies the reasoning. If $f(x)$ in particular is identically 1, each $S_k(x)$ reduces to the single term 1, and $\sigma_n(x)$ also is identically 1, for all values of n:

$$1 = \frac{1}{n\pi}\int_{-\pi}^{\pi} \frac{\sin^2 \frac{1}{2} n(t-x)}{2\sin^2 \frac{1}{2}(t-x)}\,dt = \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2} \Phi_n(u)\,du.$$

More generally, then, if $f(x)$ has M as an upper bound for its absolute value,

$$|\sigma_n(x)| \leq \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2} |f(x+2u)|\,\Phi_n(u)\,du$$
$$\leq \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2} M\Phi_n(u)\,du = M;$$

if $|f(x)| \leq M$ for all values of x, then $|\sigma_n(x)| \leq M$ for all values of x likewise, and for all values of n.

For any specified value of x, the identity in the preceding paragraph, multiplied by the quantity $f(x)$, which is constant as far as the variable of integration is concerned, states that

$$f(x) = \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2} f(x)\,\Phi_n(u)\,du,$$

whence

$$\sigma_n(x) - f(x) = \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2} [f(x+2u) - f(x)]\,\Phi_n(u)\,du.$$

Suppose now that $f(x)$ is summable over a period, and continuous for $x = x_0$. Let ε be an arbitrary positive quantity, and let $\delta \leq \pi/2$ be chosen so that $|f(x_0+2u) - f(x_0)| \leq \frac{1}{2}\varepsilon$ for $|u| \leq \delta$. Then

$$\left|\frac{1}{n\pi}\int_{-\delta}^{\delta} [f(x_0+2u) - f(x_0)]\,\Phi_n(u)\,du\right| \leq \frac{1}{n\pi}\int_{-\delta}^{\delta} \frac{1}{2}\varepsilon\,\Phi_n(u)\,du$$
$$\leq \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2} \frac{1}{2}\varepsilon\,\Phi_n(u)\,du = \frac{1}{2}\varepsilon,$$

for all values of n. On the other hand, since $\Phi_n(u) \leq 1/(\sin^2\delta)$ for $\delta \leq |u| \leq \pi/2$,

$$\left|\frac{1}{n\pi}\left(\int_{-\pi/2}^{-\delta}+\int_{\delta}^{\pi/2}\right)[f(x_0+2u)-f(x_0)]\,\Phi_n(u)\,du\right|$$
$$\leqq \frac{1}{n\pi\sin^2\delta}\int_{-\pi/2}^{\pi/2}|f(x_0+2u)-f(x_0)|\,du;$$

the last integral exists, by the hypothesis of summability, and is independent of n, so that the whole expression is less than $\frac{1}{2}\epsilon$ as soon as n is sufficiently large. Therefore $|\sigma_n(x_0)-f(x_0)|<\epsilon$, when n is sufficiently large; $\sigma_n(x_0)$ converges toward $f(x_0)$. If $f(x)$ is continuous everywhere, $\sigma_n(x)$ converges toward $f(x)$ for all values of x, and the convergence is uniform. For δ in the proof can be chosen independently of x_0, since $f(x)$ is uniformly continuous, and if M is the maximum of $|f(x)|$,

$$\int_{-\pi/2}^{\pi/2}|f(x_0+2u)-f(x_0)|\,du \leqq 2M\pi,$$

which is likewise independent of x_0.

For convergence at a point of discontinuity, let $f(x)$ be summable over a period, and have a finite jump for $x=x_0$. Let $f(x_0)$ be defined as the mean of the limits approached from the right and from the left. Let a function $\psi(x)$ be defined by the requirements that it shall be of period 2π, and linear for $x_0<x<x_0+2\pi$, and that $\psi(x_0+)=f(x_0+)$, $\psi(x_0-)=f(x_0-)$, $\psi(x_0)=f(x_0)$. For $x=x_0$, each partial sum of the Fourier series for $\psi(x)$ is exactly equal to $\psi(x_0)$, as was pointed out in substance at the beginning of the proof of Theorem II. The same is true therefore of the arithmetic means of these partial sums. But $f(x)$ is the sum of $\psi(x)$ and a function continuous for $x=x_0$; and the arithmetic mean formed for the sum of two functions is the sum of the corresponding arithmetic means. So the arithmetic mean $\sigma_n(x)$ formed for $f(x)$ converges for $x=x_0$ to the value $f(x_0)$.

The results on convergence obtained thus far may be restated in

THEOREM VII. *If $f(x)$ is a summable function of period 2π, and $\sigma_n(x)$ the arithmetic mean of the first n partial sums of*

its Fourier series, as defined above, $\sigma_n(x)$ converges toward $f(x)$ at every point where $f(x)$ is continuous, and converges toward the average of the limits approached from the right and from the left at every point where $f(x)$ has a finite jump. If $f(x)$ is continuous everywhere, the convergence is uniform throughout.

It should be mentioned as an immediate consequence of the definition of the arithmetic mean that *if $S_n(x)$ converges at any point, $\sigma_n(x)$ converges to the same value.* In fact, if $S_0, S_1, S_2, \cdots$ is any convergent sequence of numbers whatever, with limit S, and if $\sigma_n = (S_0 + S_1 + \cdots + S_{n-1})/n$, then σ_n also converges to the limit S. Let $S - S_k = R_k$, and let N be chosen so that $|R_k| < \frac{1}{2}\epsilon$ for $k > N$. Then, if $n > N$,

$$\begin{aligned} S - \sigma_n &= \frac{nS - (S_0 + S_1 + \cdots + S_{n-1})}{n} \\ &= \frac{R_0 + R_1 + \cdots + R_{n-1}}{n} \\ &= \frac{R_0 + \cdots + R_{N-1}}{n} + \frac{R_N + \cdots + R_{n-1}}{n}; \end{aligned}$$

the last fraction is less than $\frac{1}{2}\epsilon$ in absolute value for all values of $n > N$, and the preceding fraction, in which the numerator is independent of n, is less than $\frac{1}{2}\epsilon$ for n sufficiently large.

5. Degree of convergence of the first arithmetic mean

As a first hypothesis for the study of degree of convergence, let $f(x)$ satisfy everywhere the condition

$$|f(x_2) - f(x_1)| \leqq \lambda |x_2 - x_1|.$$

In the integral expression for $\sigma_n(x) - f(x)$,

$$|f(x + 2u) - f(x)| \leqq \lambda |2u|,$$

so that

$$|\sigma_n(x) - f(x)| \leqq \frac{2\lambda}{n\pi} \int_{-\pi/2}^{\pi/2} |u| \, \Phi_n(u) \, du = \frac{4\lambda}{n\pi} \int_0^{\pi/2} u \, \Phi_n(u) \, du.$$

In dealing with the last expression, it is to be remembered that $u/\sin u \leq \pi/2$ for $0<u \leq \pi/2$. Let the integrals from 0 to $1/n$ and from $1/n$ to $\pi/2$ be considered separately. In the former, let the integrand be expressed by the formula

$$u\,\Phi_n(u) = \frac{u}{\sin u}\cdot \sin nu \cdot \frac{\sin nu}{\sin u};$$

of the three factors indicated, all of which are positive for $0<u\leq 1/n$, the first does not exceed $\pi/2$, the second does not exceed 1, and the third does not exceed n. Hence

$$\int_0^{1/n} u\,\Phi_n(u)\,du \leq \frac{\pi}{2}.$$

In the integral from $1/n$ to $\pi/2$,

$$u\,\Phi_n(u) = \frac{u^2}{\sin^2 u}\cdot \sin^2 nu \cdot \frac{1}{u} \leq \frac{\pi^2}{4}\cdot\frac{1}{u},$$

and

$$\int_{1/n}^{\pi/2} u\,\Phi_n(u)\,du \leq \frac{\pi^2}{4}\int_{1/n}^{\pi/2}\frac{du}{u} = \frac{\pi^2}{4}\left(\log\frac{\pi}{2}+\log n\right).$$

It follows that the whole integral from 0 to $\pi/2$ does not exceed a constant multiple of $\log n$, and $|\sigma_n(x)-f(x)|$ does not exceed a constant multiple of $(\log n)/n$. A slightly more specific statement of the result is

THEOREM VIII. *If $f(x)$ is a function of period 2π satisfying everywhere the condition*

$$|f(x_2)-f(x_1)| \leq \lambda\,|x_2-x_1|,$$

and $\sigma_n(x)$ the corresponding arithmetic mean, then, for $n>1$.

$$|f(x)-\sigma_n(x)| \leq \frac{C_0\,\lambda\log n}{n},$$

where C_0 is an absolute constant.

Let $f(x)$ have a derivative satisfying everywhere the condition

$$|f'(x_2)-f'(x_1)| \leq \lambda\,|x_2-x_1|^{\alpha},$$

with a positive value of α. The error then can be somewhat more narrowly restricted. The factor $\Phi_n(u)$ being an even function of u, the substitution of $-u$ for u gives the integral expression for the error the alternative form

$$\sigma_n(x)-f(x) = -\frac{1}{n\pi}\int_{\pi/2}^{-\pi/2}[f(x-2u)-f(x)]\,\Phi_n(u)\,du$$
$$= \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2}[f(x-2u)-f(x)]\,\Phi_n(u)\,du,$$

and a third expression is obtained by taking the average of this and the original one:

$$\sigma_n(x)-f(x)=\frac{1}{2n\pi}\int_{\pi/2}^{\pi/2}[f(x+2u)-2f(x)+f(x-2u)]\,\Phi_n(u)\,du.$$

By the mean value theorem, together with the hypothesis now in force,

$$f(x+2u)-f(x) = 2uf'(\xi_1),$$
$$f(x-2u)-f(x) = -2uf'(\xi_2),$$
$$f(x+2u)-2f(x)+f(x-2u) = 2u[f'(\xi_1)-f'(\xi_2)],$$
$$|f(x+2u)-2f(x)+f(x-2u)| \leq 2\lambda|u|\cdot|\xi_1-\xi_2|^\alpha$$
$$\leq 2^{2\alpha+1}\lambda|u|^{\alpha+1},$$

the numbers ξ_1 and ξ_2 being in the intervals $(x, x+2u)$ and $(x-2u, x)$ respectively, so that $|\xi_1-\xi_2| \leq 4|u|$. Hence

$$|\sigma_n(x)-f(x)| \leq \frac{2^{2\alpha}\lambda}{n\pi}\int_{-\pi/2}^{\pi/2}|u|^{\alpha+1}\,\Phi_n(u)\,du$$
$$= \frac{2^{2\alpha+1}\lambda}{n\pi}\int_0^{\pi/2}u^{\alpha+1}\,\Phi_n(u)\,du.$$

But

$$\Phi_n(u) \leq \frac{1}{\sin^2 u} \leq \frac{\pi^2}{4}\cdot\frac{1}{u^2}$$

in the interval of integration, so that

$$|\sigma_n(x)-f(x)| \leq \frac{2^{2\alpha-1}\lambda\pi}{n}\int_0^{\pi/2}\frac{du}{u^{1-\alpha}},$$

which does not exceed a constant multiple of $1/n$, the last integral being convergent. In more formal statement:

THEOREM IX. *If $f(x)$ is a function of period 2π having a first derivative which satisfies everywhere the condition*

$$|f'(x_2) - f'(x_1)| \leqq \lambda\, |x_2 - x_1|^{\alpha}$$

with $\alpha > 0$, and if $\sigma_n(x)$ is the corresponding arithmetic mean, then

$$|f(x) - \sigma_n(x)| \leqq \frac{C_\alpha \lambda}{n},$$

where C_α is a constant depending only on α.

No higher order of approximation would be obtained by supposing $f(x)$ provided with additional derivatives. In fact, the arithmetic mean corresponding to the analytic function $f(x) = \cos x$ is $\sigma_n(x) = [(n-1)/n] \cos x$, and the error is actually of the order of $1/n$.

Let $f(x)$ be an arbitrary continuous function of period 2π, with modulus of continuity $\omega(\delta)$. The theorem to be obtained in this connection is perhaps of secondary interest, because a closer result is given for an important class of cases by the theorem following it; but it also covers cases not admitted by the hypothesis of the later theorem, and so is not entirely superfluous. The proof is an adaptation of that of Theorem II in Chapter I. Let $\varphi(x)$ be a continuous function of period 2π which is equal to $f(x)$ for a set of values of x dividing a period into n equal parts, and is linear between successive points of this set. This $\varphi(x)$ satisfies the hypothesis of Theorem VIII, with $\lambda = [\omega(2\pi/n)]/(2\pi/n)$, and is represented by the corresponding arithmetic mean with an error not exceeding a constant multiple of $\omega(2\pi/n) \log n$. The absolute value of the mean corresponding to the difference $f(x) - \varphi(x)$ does not exceed the maximum of the absolute value of the difference itself, which is not greater than $2\omega(2\pi/n)$, and the error of this mean can not be greater than $4\omega(2\pi/n)$. Hence:

THEOREM X. *If $f(x)$ is a continuous function of period 2π, with modulus of continuity $\omega(\delta)$, and $\sigma_n(x)$ the corresponding arithmetic mean, then, for $n > 1$,*

$$|f(x) - \sigma_n(x)| \leq C_0' \omega\left(\frac{2\pi}{n}\right) \log n,$$

where C_0' is an absolute constant.

Let $f(x)$ be subjected to the hypothesis that

$$|f(x_2) - f(x_1)| \leq \lambda |x_2 - x_1|^\alpha,$$

with $0 < \alpha < 1$. From the original integral expression for the error,

$$|\sigma_n(x) - f(x)| \leq \frac{\lambda}{n\pi} \int_{-\pi/2}^{\pi/2} |2u|^\alpha \Phi_n(u)\, du$$
$$= \frac{2^{\alpha+1}\lambda}{n\pi} \int_0^{\pi/2} u^\alpha \Phi_n(u)\, du.$$

Let the interval of integration again be considered in two parts, from 0 to $1/n$ and from $1/n$ to $\pi/2$. In the former interval,

$$u^\alpha \Phi_n(u) = \frac{u^\alpha}{\sin^\alpha u} \cdot \sin^\alpha nu \cdot \left(\frac{\sin nu}{\sin u}\right)^{2-\alpha} \leq \left(\frac{\pi}{2}\right)^\alpha \cdot 1 \cdot n^{2-\alpha},$$

so that

$$\int_0^{1/n} u^\alpha \Phi_n(u)\, du < \left(\frac{\pi}{2}\right)^\alpha n^{1-\alpha}.$$

From $1/n$ to $\pi/2$,

$$u^\alpha \Phi_n(u) \leq \frac{u^\alpha}{\sin^2 u} \leq \frac{\pi^2}{4} \cdot \frac{1}{u^{2-\alpha}},$$

$$\int_{1/n}^{\pi/2} u^\alpha \Phi_n(u)\, du \leq \frac{\pi^2}{4} \int_{1/n}^{\pi/2} \frac{du}{u^{2-\alpha}} < \frac{\pi^2}{4} \int_{1/n}^{\infty} \frac{du}{u^{2-\alpha}} = \frac{\pi^2 n^{1-\alpha}}{4(1-\alpha)}.$$

Substitution of these inequalities in the formula for the error gives

THEOREM XI. *If $f(x)$ is a function of period 2π satisfying everywhere the condition*

$$|f(x_2) - f(x_1)| \leq \lambda |x_2 - x_1|^\alpha,$$

with $0 < \alpha < 1$, and if $\sigma_n(x)$ is the corresponding arithmetic mean, then

$$|f(x) - \sigma_n(x)| \leq \frac{C_\alpha'' \lambda}{n^\alpha},$$

where C_α'' is a constant depending only on α.

Two outstanding facts with regard to the arithmetic mean, as compared with the simple partial sum of the Fourier series, are that the former converges for every continuous function, but does not reproduce identically a function which is itself a finite trigonometric sum. Consideration of Theorems XI, VIII, and IX, in the order named, throws some further light on the ability of the arithmetic mean to adapt itself to irregularities in the function represented, and its inability to avail itself of an exceptional degree of regularity. The hypotheses of the three theorems imply that

$$|f(x+v)-2f(x)+f(x-v)| \leqq \lambda_1 |v|^\alpha$$

with a constant λ_1 in each case, and values of α successively less than 1, equal to 1, and greater than 1 (the present α, in the case of Theorem IX, taking the place of the number previously denoted by $\alpha+1$). The upper bounds obtained for the error of the arithmetic mean have the orders respectively of $1/n^\alpha$, $(\log n)/n^\alpha$, and $n^{\alpha-1}/n^\alpha$. The corresponding upper bound for the error of the simple partial sum of the Fourier series is of the order of $(\log n)/n^\alpha$ in each case.

For a concluding theorem with regard to the arithmetic mean, let $f(x)$ be of period 2π, summable over a period, and identically zero for $\alpha-\eta \leqq x \leqq \beta+\eta$. If x has any value in (α, β), $f(x+2u)$ is identically zero for $-\frac{1}{2}\eta \leqq u \leqq \frac{1}{2}\eta$. Hence

$$\sigma_n(x) = \frac{1}{n\pi}\left(\int_{-\pi/2}^{-\eta/2} + \int_{\eta/2}^{\pi/2}\right) f(x+2u)\, \Phi_n(u)\, du.$$

Since $\Phi_n(u) \leqq 1/(\sin^2 \frac{1}{2}\eta)$ for $\frac{1}{2}\eta \leqq |u| \leqq \frac{1}{2}\pi$, it follows that

$$|\sigma_n(x)| \leqq \frac{1}{n\pi \sin^2 \frac{1}{2}\eta}\left(\int_{-\pi/2}^{-\eta/2} + \int_{\eta/2}^{\pi/2}\right) |f(x+2u)|\, du$$

$$\leqq \frac{1}{n\pi \sin^2 \frac{1}{2}\eta}\int_{-\pi/2}^{\pi/2} |f(x+2u)|\, du.$$

In terms of the variable $t = x+2u$,

$$\int_{-\pi/2}^{\pi/2} |f(x+2u)|\, du = \frac{1}{2}\int_{-\pi}^{\pi} |f(t)|\, dt,$$

the limits of integration on the right having been adjusted by use of the frequently applied observation that the integral of a periodic function over a period is independent of the position of the initial point of the period. If the last integral, which is independent of x, is denoted by J,

$$|\sigma_n(x)| \leq \frac{J}{2n\pi \sin^2 \frac{1}{2}\eta}.$$

The conclusion is

THEOREM XII. *If $f(x)$ is a function of period 2π which is identically zero for $\alpha - \eta \leq x \leq \beta + \eta$, and summable over a period, the integral of its absolute value over a period being J, and if $\sigma_n(x)$ is the corresponding arithmetic mean, then, for $\alpha \leq x \leq \beta$,*

$$|\sigma_n(x)| \leq \frac{C_\eta J}{n},$$

where C_η is a constant depending only on η.

Taken with the last assertion in Theorem VII, this shows that if $f(x)$ is continuous for $\alpha - \eta \leq x \leq \beta + \eta$, and summable over a period, the arithmetic mean associated with it converges uniformly for $\alpha \leq x \leq \beta$. The results obtained by combining Theorem XII with Theorems VIII–XI need not be enumerated at length.

6. Convergence of Legendre series under hypothesis of continuity over a part of the interval

The theory set forth in the early part of the chapter can be carried over in some measure to the case of Legendre series, the discussion being kept on the same elementary level which was maintained in the treatment of these series in Chapter I, to the extent that no use is made of an asymptotic formula for the Legendre polynomials. The next paragraphs will be devoted to a presentation of the analogies that work out most readily, though it will be seen that the treatment is left incomplete in several particulars.

Let $f(x)$ be summable together with its square for $-1 \leq x \leq 1$, and let a_k be the coefficient of $X_k(x)$ in its Legendre series,

$$a_k = \frac{2k+1}{2}\int_{-1}^{1} f(t)\, X_k(t)\, dt.$$

Let

$$S_n(x) = a_0 X_0(x) + a_1 X_1(x) + \cdots + a_n X_n(x),$$

as before. By reference to the property of orthogonality of the polynomials $X_k(x)$, together with the definition of a_k and the fact that the integral of $[X_k(x)]^2$ over the interval $(-1, 1)$ is $2/(2k+1)$, it is seen that

$$\int_{-1}^{1}[f(x) - S_n(x)]^2\, dx = \int_{-1}^{1}[f(x)]^2\, dx - \sum_{k=0}^{n}\frac{2}{2k+1}\, a_k^2.$$

As the quantity on the left-hand side can not be negative, the sum on the right is bounded for all values of n, the infinite series obtained by letting n become infinite is convergent, and

$$\lim_{k=\infty} a_k/(2k+1)^{1/2} = 0,$$

or, as equivalent statements,

$$\lim_{k=\infty} a_k/k^{1/2} = 0, \qquad \lim_{k=\infty} k^{1/2}\int_{-1}^{1} f(t)\, X_k(t)\, dt = 0.$$

The parallelism with the case of Fourier series becomes clearer if the polynomials $X_k(x)$ are replaced by the normalized sequence $\{[(2k+1)/2]^{1/2}\, X_k(x)\}$, or, more superficially, if it is considered that the magnitude of the coefficient is less significant than the magnitude of the general term of the series: it was seen in Chapter I that $|X_k(x)|$ does not exceed a constant divided by $k^{1/2}$, when x stays away from the ends of the interval, and consequently

$$\lim_{k=\infty} a_k X_k(x) = 0$$

uniformly throughout any closed interval interior to $(-1, 1)$.

Let $f(x)$ again be summable together with its square, and suppose now that it vanishes identically for $x_0 - \eta \leqq x \leqq x_0 + \eta$, the interval $(x_0 - \eta,\, x_0 + \eta)$ being contained in $(-1, 1)$. The value of $S_n(x_0)$ can be expressed in the form

$$S_n(x_0) = \frac{n+1}{2} X_{n+1}(x_0) \int_{-1}^{1} \frac{f(t)}{x_0 - t} X_n(t)\,dt$$
$$- \frac{n+1}{2} X_n(x_0) \int_{-1}^{1} \frac{f(t)}{x_0 - t} X_{n+1}(t)\,dt.$$

Each of the quantities $X_{n+1}(x_0)$, $X_n(x_0)$ is less than a constant divided by $n^{1/2}$, and the factor in front of each integral is therefore less than a constant multiple of $n^{1/2}$; the quantity $1/(x_0 - t)$ is bounded over the range where $f(t)$ is different from zero, and $f(t)/(x_0 - t)$ therefore is summable together with its square; consequently, by the preceding paragraph, the product of either integral by a quantity of the order $n^{1/2}$ approaches zero. In other words, $S_n(x_0)$ *converges to the value* 0. If two functions $f(x)$, $\varphi(x)$ are identical for $x_0 - \eta \leqq x \leqq x_0 + \eta$, the Legendre series for their difference converges toward zero at the point x_0, and the series for $f(x)$ and $\varphi(x)$ themselves converge or diverge together at that point: *the convergence of the Legendre series for a given function at an interior point of the interval* $(-1, 1)$ *depends only on the behavior of the function in the neighborhood of the point.*

As a preliminary to a discussion of uniform convergence, let $f(x)$ be any summable function with summable square over the interval $(-1, 1)$, and let

$$A_k(y) = \int_{-1}^{y} f(t)\, X_k(t)\,dt.$$

It is to be shown that $k^{1/2} A_k(y)$ approaches zero *uniformly for* $-1 \leqq y \leqq 1$. The proof is largely a repetition of the corresponding argument for the case of Fourier series, but there are differences of detail which are not altogether trivial. For any fixed value of y, $A_k(y)$ is the same as

$$\int_{-1}^{1} f_1(t)\, X_k(t)\,dt,$$

if $f_1(t)$ is a function equal to $f(t)$ for $-1 \leqq t \leqq y$, and vanishing for $y < t \leqq 1$, and consequently $k^{1/2} A_k(y)$ approaches

zero for the particular value of y in question. Let ε be any positive quantity. Let

$$A(y) = \int_{-1}^{y} [f(t)]^2 \, dt.$$

This function is continuous, and therefore uniformly continuous, for $-1 \leqq y \leqq 1$. Let δ be chosen so that

$$|A(y_2) - A(y_1)| < \left(\frac{1}{2}\varepsilon\right)^2$$

whenever $|y_2 - y_1| < \delta$, y_1 and y_2 being points of the interval $(-1, 1)$. Let N be the smallest integer such that $2/N < \delta$, and let $z_j = -1 + (2j/N)$, $j = 0, 1, \cdots, N$. The points z_j being finite in number, there is a k' such that $k^{1/2}|A_k(z_j)| < \frac{1}{2}\varepsilon$ for all the values of j in question, whenever $k \geqq k'$. For any value of y in $(-1, 1)$, let j be that one of the numbers $0, \cdots, N$ for which $0 \leqq y - z_j < 2/N < \delta$, and let $Z = z_j$. Consider the difference

$$A_k(y) - A_k(Z) = \int_{Z}^{y} f(t) \, X_k(t) \, dt.$$

By Schwarz's inequality,

$$\begin{aligned}
\left[\int_{Z}^{y} f(t) \, X_k(t) \, dt\right]^2 &\leqq \int_{Z}^{y} [f(t)]^2 \, dt \int_{Z}^{y} [X_k(t)]^2 \, dt \\
&= [A(y) - A(Z)] \int_{Z}^{y} [X_k(t)]^2 \, dt \\
&\leqq [A(y) - A(Z)] \int_{-1}^{1} [X_k(t)]^2 \, dt \\
&\leqq \left(\frac{1}{2}\varepsilon\right)^2 \cdot 2/(2k+1) < \left(\frac{1}{2}\varepsilon\right)^2 / k.
\end{aligned}$$

Consequently

$$k^{1/2}|A_k(y) - A_k(Z)| < \frac{1}{2}\varepsilon.$$

This is true for any positive value of k. On the other hand, if $k \geqq k'$, $k^{1/2}|A_k(Z)| < \frac{1}{2}\varepsilon$. For such values of k, therefore, $k^{1/2}|A_k(y)| < \varepsilon$; and k' is independent of y. Let ε_k be the maximum of $|A_k(y)|$ for $-1 \leqq y \leqq 1$. The conclusion is that $\lim_{k \to \infty} k^{1/2} \varepsilon_k = 0$.

Now let $f(x)$ be summable together with its square over the interval $(-1, 1)$, and identically zero for $\alpha - \eta \leqq x \leqq \beta + \eta$, where $-1 < \alpha - \eta < \beta + \eta < 1$. Consider the expression

$$\int_{-1}^{1} \frac{f(t)\, X_n(t)}{x - t}\, dt = \int_{-1}^{1} \frac{A_n'(t)}{x - t}\, dt,$$

on the assumption that x has a value belonging to the interval $\alpha \leqq x \leqq \beta$; the function $A_n(t)$ has almost everywhere a derivative equal to $f(t)\, X_n(t)$. Let $R(u)$ be a function equal to $1/u$ for $|u| \geqq \eta$, and so defined for $|u| < \eta$ as to have a continuous derivative everywhere. Then, for any x in (α, β), $R(x - t)$ is the same as $1/(x - t)$ at all points of the interval $-1 \leqq t \leqq 1$ where $f(t)$ is different from zero, and the integral above is equal to

$$\int_{-1}^{1} A_n'(t)\, R(x - t)\, dt.$$

The product $A_n(t)\, R(x - t)$ has almost everywhere a derivative with regard to t, equal to $A_n'(t)\, R(x - t) - A_n(t)\, R'(x - t)$, and its increment over an interval is equal to the integral of this derivative, so that

$$\int_{-1}^{1} A_n'(t)\, R(x - t)\, dt$$
$$= [A_n(t)\, R(x - t)]_{-1}^{1} + \int_{-1}^{1} A_n(t)\, R'(x - t)\, dt.$$

Both $R(u)$ and $R'(u)$ are continuous for all real values of u, and approach zero as u becomes infinite in either direction. Let M_0 be the maximum of $|R(u)|$, and M_1 the maximum of $|R'(u)|$; these numbers are of course independent of x. In the right-hand member of the last equality, $A_n(-1) = 0$, $|A_n(1)| \leqq \epsilon_n$, $|R(x - 1)| \leqq M_0$, while $|A_n(t)| \leqq \epsilon_n$, $|R'(x - t)| \leqq M_1$, over the whole interval of integration. Consequently

$$\left| \int_{-1}^{1} A_n'(t)\, R(x - t)\, dt \right| \leqq M_0\, \epsilon_n + 2\, M_1\, \epsilon_n.$$

Since $\lim_{n=\infty} n^{1/2}\, \epsilon_n = 0$,

$$n^{1/2}\int_{-1}^{1}\frac{f(t)\,X_n(t)}{x-t}\,dt$$

approaches zero uniformly for $\alpha \leqq x \leqq \beta$. The same conclusion holds if $X_n(t)$ is replaced by $X_{n+1}(t)$. On the other hand, $n^{1/2}X_{n+1}(x)$ and $n^{1/2}X_n(x)$ are uniformly bounded over (α, β). Hence, on assembling the constituent parts of the integral expression for $S_n(x)$, it is seen that $S_n(x)$ approaches zero uniformly over the interval in question:

THEOREM XIII. *If $f(x)$ is summable together with its square over the interval $(-1, 1)$, and identically zero over an interval $\alpha-\eta \leqq x \leqq \beta+\eta$ contained in $(-1, 1)$, and if $S_n(x)$ is the partial sum of the Legendre series for $f(x)$, then $S_n(x)$ converges uniformly toward 0 for $\alpha \leqq x \leqq \beta$.*

With the corollary of Theorem XI in Chapter I, this yields at once the further

COROLLARY. *If $f(x)$ is summable together with its square over the interval $(-1, 1)$, and continuous over the interval $\alpha-\eta \leqq x \leqq \beta+\eta$, with a modulus of continuity $\omega(\delta)$ such that $\lim_{\delta=0}\omega(\delta)\log\delta = 0$, the Legendre series for $f(x)$ converges uniformly to the value $f(x)$ for $\alpha \leqq x \leqq \beta$.*

7. Degree of convergence of Legendre series under hypotheses involving limited variation

It remains to consider questions of degree of convergence. The discussion will be based on the properties of the Legendre polynomials already used, together with the identity

$$X_n(x) = \frac{1}{2n+1}\left[X'_{n+1}(x) - X'_{n-1}(x)\right].$$

Let $f(x)$ be a function of limited variation for $-1 \leqq x \leqq 1$, its total variation being V. Let $\varphi_1(x)$ and $\varphi_2(x)$ denote its positive and negative variations, measured from the left-hand end of the interval, so that $f(x) = f(-1) + \varphi_1(x) - \varphi_2(x)$, $\varphi_1(-1) = \varphi_2(-1) = 0$. By the second law of the mean,

$$\int_{-1}^{1}\varphi_1(x)\,X_n(x)\,dx = \varphi_1(1)\int_{\xi}^{1}X_n(x)\,dx$$

for some value of ξ in $(-1, 1)$. But as a consequence of the identity in the preceding paragraph, with the observation that $X_{n+1}(1) = X_{n-1}(1) = 1$,

$$\int_\xi^1 X_n(x)\,dx = \frac{1}{2n+1}\left[X_{n+1}(x) - X_{n-1}(x)\right]_\xi^1$$
$$= \frac{1}{2n+1}\left[X_{n-1}(\xi) - X_{n+1}(\xi)\right].$$

For a sufficiently close estimate of the magnitude of the difference in the last bracket, there is occasion to go back to the identity

$$X_n(x) = \frac{1}{\pi}\int_0^\pi \left[x + i(1-x^2)^{1/2}\cos\varphi\right]^n d\varphi.$$

If this formula is used to express the difference $X_{n-1}(x) - X_{n+1}(x)$ by means of a single integral, the integrand is

$$\left[x + i(1-x^2)^{1/2}\cos\varphi\right]^{n-1}\left\{1 - \left[x + i(1-x^2)^{1/2}\cos\varphi\right]^2\right\}$$
$$= \left[x + i(1-x^2)^{1/2}\cos\varphi\right]^{n-1}(1-x^2)^{1/2}\left[(1-x^2)^{1/2}(1+\cos^2\varphi) - 2ix\cos\varphi\right].$$

For x in $(-1, 1)$, the absolute value of the last bracket can not exceed 4, and hence

$$|X_{n-1}(x) - X_{n+1}(x)| < \frac{4}{\pi}(1-x^2)^{1/2}\int_0^\pi |x + i(1-x^2)^{1/2}\cos\varphi|^{n-1} d\varphi.$$

On repeating, with slight changes of notation, the reasoning which led to inequalities for the Legendre polynomials in Chapter I, it is found that

$$\int_0^\pi |x + i(1-x^2)^{1/2}\cos\varphi|^{n-1} d\varphi = 2\int_0^{\pi/2}\left[x^2 + (1-x^2)\cos^2\varphi\right]^{(n-1)/2} d\varphi,$$
$$x^2 + (1-x^2)\cos^2\varphi \leqq e^{-z^2\varphi^2}$$

for the values of the variables that come into consideration, if $z = (2/\pi)(1-x^2)^{1/2}$, and

$$\int_0^{\pi/2} e^{-(n-1)z^2\varphi^2/2}\,d\varphi \leqq \frac{2^{1/2}}{(n-1)^{1/2}z}\int_0^\infty e^{-u^2}\,du$$
$$= \frac{\pi}{[2(n-1)(1-x^2)]^{1/2}}\int_0^\infty e^{-u^2}\,du.$$

With the observation that $n-1 \geqq \frac{1}{2}n$ for $n \geqq 2$, the preceding inequalities may be summarized by saying that there is an absolute constant c such that

$$|X_{n-1}(x) - X_{n+1}(x)| \leqq c/n^{1/2}$$

for $n \geqq 1$ and for $-1 \leqq x \leqq 1$; it is clear that the case $n=1$ can be included in the final statement. (Strictly speaking, of course, the result is obtained directly for $|x| < 1$, and extended to the ends of the interval by continuity.) The essential point is that there is no factor $(1-x^2)^{1/2}$ in the denominator on the right, as there was in the corresponding inequality for a single Legendre polynomial; its absence is due to the factor $(1-x^2)^{1/2}$ which comes out before the integral sign in the expression for the difference.

Applied to the problem in hand, the inequality just obtained shows that

$$\left| \int_{-1}^{1} \varphi_1(x)\, X_n(x)\, dx \right| \leqq \frac{\varphi_1(1)}{2n+1} \cdot \frac{c}{n^{1/2}} \leqq \frac{c\,\varphi_1(1)}{2n^{3/2}}.$$

Similarly, the magnitude of the integral formed with $\varphi_2(x)$ in place of $\varphi_1(x)$ does not exceed $c\varphi_2(1)/(2n^{3/2})$, while for $n > 1$ the integral corresponding to the constant $f(-1)$ is zero. Consequently, as $\varphi_1(1) + \varphi_2(1) = V$,

$$\left| \int_{-1}^{1} f(x)\, X_n(x)\, dx \right| \leqq \frac{cV}{2n^{3/2}}.$$

The analogue of Theorem III is

THEOREM XIV. *If $f(x)$ is of limited variation for $-1 \leqq x \leqq 1$, its total variation being V, and if a_n is the coefficient of $X_n(x)$ in its Legendre series, then, for $n > 0$,*

$$|a_n| \leqq \frac{R_0 V}{n^{1/2}},$$

where R_0 is an absolute constant, so that

$$|a_n X_n(x)| \leqq \frac{R_\eta V}{n}$$

for $-1+\eta \leqq x \leqq 1-\eta$, the constant R_η depending only on η.

Now let it be assumed that $f(x)$ has a first derivative with limited variation, and let the total variation of $f'(x)$ for $-1 \leqq x \leqq 1$ be denoted by V. (In more general terms, it would be sufficient that $f(x)$ be the integral of a function of limited variation, and not necessarily provided with a unique derivative at every point.) For the moment, let

$$Y_n(x) = \int_{-1}^{x} X_n(t)\,dt.$$

By integration by parts,

$$\int_{-1}^{1} f(x)\,X_n(x)\,dx = \int_{-1}^{1} f(x)\,Y_n'(x)\,dx$$
$$= [f(x)\,Y_n(x)]_{-1}^{1} - \int_{-1}^{1} f'(x)\,Y_n(x)\,dx.$$

In the last member, $Y_n(-1) = 0$, and for $n \geqq 1$, $Y_n(1) = 0$ also. Furthermore, by the identity already employed,

$$Y_n(x) = \frac{1}{2n+1}\,[X_{n+1}(x) - X_{n-1}(x)];$$

the terms coming from the lower limit of integration cancel each other, since $X_{n+1}(-1) = X_{n-1}(-1) = (-1)^{n+1}$. Hence. with the assumption that $n \geqq 1$,

$$\int_{-1}^{1} f(x)\,X_n(x)\,dx = \frac{1}{2n+1}\int_{-1}^{1} f'(x)\,[X_{n-1}(x) - X_{n+1}(x)]\,dx.$$

If the integral on the right is taken as the difference of two integrals, Theorem XIV, or, more directly, the inequality immediately preceding the statement of that theorem, can be applied to each. It is to be noted once more that $1/(n-1) \leqq 2/n$ for $n > 1$. The resulting general formulation, however, does *not* hold for $n = 1$, as may be seen by taking $f(x) \equiv Gx$, in which case $V = 0$, while G may be arbitrarily large; the conclusion is essentially restricted to values of $n \geqq 2$. Induction based on repetition of the process of integration by parts leads to a conclusion which may be stated as

COROLLARY I. *If $f(x)$ has a pth derivative of limited variation, $p \geqq 0$, and if V is the total variation of $f^{(p)}(x)$ for $-1 \leqq x \leqq 1$, then, for $n > p$,*

$$|a_n| \leqq \frac{R_p V}{n^{p+(1/2)}},$$

where R_p depends only on p, and

$$|a_n X_n(x)| \leqq \frac{R_{p\eta} V}{n^{p+1}}$$

for $-1+\eta \leqq x \leqq 1-\eta$, where $R_{p\eta}$ depends on p and on η, but not on x or on n.

Suppose $f(x)$ is of limited variation for $-1 \leqq x \leqq 1$, with total variation V, and identically zero for $\alpha-\eta \leqq x \leqq \beta+\eta$. Let $R(u)$, as before, be equal to $1/u$ for $|u| \geqq \eta$, and defined in some way for $|u| < \eta$ so as to have a continuous derivative everywhere. If M_1 is the maximum of $|R'(u)|$, the total variation of $R(u)$ over any interval of length 2 does not exceed $2M_1$. The quotient $f(t)/(x-t)$ is identical with $f(t)\,R(x-t)$, if x has any value belonging to the interval $\alpha \leqq x \leqq \beta$. It is of limited variation, regarded as a function of t, and its total variation has an upper bound which may be expressed as the product of V by a quantity depending only on η, and, in particular, independent of x. So Theorem XIV may be applied to the function $f(t)\,R(x-t)$, and the result substituted in the integral expression for $S_n(x)$, to prove

COROLLARY II. *If $f(x)$ is of limited variation for $-1 \leqq x \leqq 1$, with total variation V, and identically zero throughout an interval $\alpha-\eta \leqq x \leqq \beta+\eta$ contained in $(-1, 1)$, and if $S_n(x)$ is the partial sum of its Legendre series, then, for $n \geqq 1$ and $\alpha \leqq x \leqq \beta$,*

$$|S_n(x)| \leqq \frac{C'_\eta V}{n},$$

where C'_η depends only on η.

Further consequences can be deduced substantially as in the case of Fourier series. If $f(x)$ has a bounded derivative for $-1 \leqq x \leqq 1$, it comes under the hypothesis of the Corollary of Theorem XI in Chapter I, and its Legendre series converges to the value $f(x)$ for $-1 < x < 1$. If the series converges uniformly for $-1 \leqq x \leqq 1$, its sum agrees in value with $f(x)$ at the ends of the interval also, by continuity.

So Corollary I above yields information as to the order of magnitude of the error, under the hypotheses indicated. A result may be stated for the entire interval $(-1, 1)$, as well as for an interior interval, since $|a_n X_n(x)| \leq |a_n|$ for $-1 \leq x \leq 1$:

THEOREM XV. *If $f(x)$ has a pth derivative with limited variation, $p \geq 1$, if V is the total variation of $f^{(p)}(x)$ for $-1 \leq x \leq 1$, and if $S_n(x)$ is the partial sum of the Legendre series for $f(x)$, then, for $n \geq p$,*

$$|f(x) - S_n(x)| \leq \frac{Q'_p V}{n^{p-(1/2)}}$$

for $-1 \leq x \leq 1$, where Q'_p depends only on p, and

$$|f(x) - S_n(x)| \leq \frac{Q_{p\eta} V}{n^p}$$

for $-1+\eta \leq x \leq 1-\eta$, where $Q_{p\eta}$ depends only on p and on η.

In conclusion, the following consequences of Corollary II, taken first with the corollary of Theorem XI in Chapter I and then with the Theorem XV just formulated, may be stated without further comment:

THEOREM XVIa. *If $f(x)$ is continuous with modulus of continuity $\omega(\delta)$ throughout an interval $\alpha - \eta \leq x \leq \beta + \eta$ contained in $(-1, 1)$, where $\omega(\delta) > 0$ for $\delta > 0$, and of limited variation over the rest of the interval $(-1, 1)$, then*

$$|f(x) - S_n(x)| \leq c\,\omega(2/n) \log n$$

for $\alpha \leq x \leq \beta$, if n is large enough so that $\omega(2/n)$ has a meaning, $S_n(x)$ being the partial sum of the Legendre series for $f(x)$, and c a constant depending neither on x nor on n.

THEOREM XVIb. *If $f(x)$ is of limited variation for $-1 \leq x \leq 1$, and has a first derivative of limited variation throughout an interval $\alpha - \eta \leq x \leq \beta + \eta$ contained in $(-1, 1)$, then*

$$|f(x) - S_n(x)| \leq \frac{c}{n}$$

for $\alpha \leq x \leq \beta$, where $S_n(x)$ is the partial sum of the Legendre series for $f(x)$, and c is a constant depending neither on x nor on n.

CHAPTER III

THE PRINCIPLE OF LEAST SQUARES AND ITS GENERALIZATIONS

1. Convergence of trigonometric approximation as related to integral of square of error

In the discussion of Fourier series hitherto, scarcely any mention has been made of one of their most remarkable and important properties, namely their relation to a problem of least squares. The property in question is not peculiar to Fourier series, but is of much wider significance.

Let $p_0(x), p_1(x), p_2(x), \cdots$, be a sequence of normalized orthogonal functions over an interval (a, b), that is, a sequence of functions satisfying the conditions

$$\int_a^b p_j(x)\, p_k(x)\, dx = 0 \quad (j \neq k), \qquad \int_a^b [p_k(x)]^2 dx = 1.$$

Let $f(x)$ be another function defined over the same interval. The functions $p_k(x)$ and $f(x)$ may be continuous, or, more generally, bounded and integrable in the sense of Riemann, or merely summable together with their squares. The problem of least squares in question is that of determining a set of coefficients $c_0, \cdots, c_n$, for a given value of n, so that when

$$\varphi(x) = c_0 p_0(x) + c_1 p_1(x) + \cdots + c_n p_n(x)$$

the integral

$$\int_a^b [f(x) - \varphi(x)]^2 dx,$$

regarded as a function of its coefficients, shall be a minimum. The following theorem is well known:

THEOREM I. *The integral has a minimum value, for the attainment of which it is necessary and sufficient that*

$$c_k = \int_a^b f(x)\, p_k(x)\, dx, \qquad k = 0, 1, \cdots, n.$$

It is to be noticed at the outset that no function of the form $\varphi(x)$ can vanish identically over the interval unless all the coefficients c_k are zero; in other words, the conditions on the p's insure that they are linearly independent. For

$$\int_a^b [\varphi(x)]^2 \, dx = c_0^2 + c_1^2 + \cdots + c_n^2,$$

and the identical vanishing of φ would require the vanishing of the sum of the squares. More definitely, it is recognized that the value of the integral on the left must be positive, if the c's are not all zero. If the p's are continuous, this means that there must be one or more intervals throughout which φ remains different from zero; if they are not so restricted, it means in any case that φ must be different from zero over a set of positive measure.

To prove the necessity of the condition, suppose that $\varphi(x)$ is a function of the form $\sum c_k p_k(x)$, in which, for a particular ndex $k = m$,

$$c_m \neq \int_a^b f(x) \, p_m(x) \, dx.$$

Let

$$\begin{aligned} \psi(x) &= c_0 p_0(x) + c_1 p_1(x) + \cdots + (c_m + h) p_m(x) + \cdots + c_n p_n(x) \\ &= \varphi(x) + h \, p_m(x), \end{aligned}$$

where h is a constant, the value of which will be specified presently. Then $f(x) - \psi(x) = f(x) - \varphi(x) - h p_m(x)$, and

$$\begin{aligned} \int_a^b [f(x) - \psi(x)]^2 \, dx &= \int_a^b [f(x) - \varphi(x)]^2 \, dx \\ &\quad - 2h \int_a^b [f(x) - \varphi(x)] \, p_m(x) \, dx + h^2 \int_a^b [p_m(x)]^2 \, dx. \end{aligned}$$

The last integral in this equation is equal to 1; if the others are denoted by J, R, S respectively,

$$J = R - 2hS + h^2.$$

But $\int_a^b \varphi(x) \, p_m(x) \, dx = c_m$, and

$$S = \int_a^b [f(x) - \varphi(x)] \, p_m(x) \, dx = \int_a^b f(x) p_m(x) \, dx - c_m \neq 0,$$

by hypothesis. Let h be taken equal to S; then

$$J = R - S^2 < R,$$

and the value R corresponding to the function φ is not the smallest value of the integral, since the function ψ gives a smaller value.

To show that the condition is sufficient, suppose now that $\varphi = \sum c_k p_k$ with $c_k = \int_a^b f(x)\, p_k(x)\, dx$ for $k = 0, 1, \cdots, n$, and let

$$\psi(x) = c_0' p_0(x) + c_1' p_1(x) + \cdots + c_n' p_n(x),$$

where at least one c_k' is different from the corresponding c_k. Then $\varphi(x) - \psi(x)$ is a linear combination of $p_0, \cdots, p_n$, with coefficients which are not all zero, and

$$V = \int_a^b [\varphi(x) - \psi(x)]^2\, dx > 0.$$

On the other hand,

$$\int_a^b f(x)\, p_k(x)\, dx = c_k = \int_a^b \varphi(x)\, p_k(x)\, dx,$$

$$\int_a^b [f(x) - \varphi(x)]\, p_k(x)\, dx = 0, \qquad k = 0, 1, \cdots, n,$$

so that if $\chi(x) = \sum c_k'' p_k(x)$ is any linear combination of $p_0, \cdots, p_n$,

$$\int_a^b [f(x) - \varphi(x)]\, \chi(x)\, dx$$

$$= \sum_{k=0}^{n} c_k'' \int_a^b [f(x) - \varphi(x)]\, p_k(x)\, dx = 0.$$

In particular,

$$U = \int_a^b [f(x) - \varphi(x)]\, [\varphi(x) - \psi(x)]\, dx = 0.$$

Hence, by the resolution $f - \psi = (f - \varphi) + (\varphi - \psi)$,

$$\int_a^b [f(x) - \psi(x)]^2\, dx = \int_a^b [f(x) - \varphi(x)]^2\, dx + 2U + V$$

$$> \int_a^b [f(x) - \varphi(x)]^2\, dx,$$

which means that φ actually gives the integral a smaller value than any other linear combination of $p_0, \cdots, p_n$.

As the most important of specific cases, $p_0, p_1, p_2, \cdots$ may be the sequence of functions

$$\frac{1}{(2\pi)^{1/2}}, \quad \frac{\cos x}{\pi^{1/2}}, \quad \frac{\sin x}{\pi^{1/2}}, \quad \frac{\cos 2x}{\pi^{1/2}}, \quad \frac{\sin 2x}{\pi^{1/2}}, \cdots.$$

If n above is replaced by $n' = 2n + 1$, $\varphi(x)$ is a trigonometric sum $T_n(x)$ of the nth order. A periodic function $f(x)$ being given, the condition that the integral

$$\int_{-\pi}^{\pi} [f(x) - T_n(x)]^2 \, dx$$

shall be a minimum, is that the coefficient of $\cos kx/\pi^{1/2}$ be

$$\int_{-\pi}^{\pi} f(t) \cdot \frac{\cos kt}{\pi^{1/2}} \, dt$$

for $k > 0$, with a corresponding determination of the constant term and the sine terms, and this is equivalent to saying that $T_n(x)$ shall be the partial sum of the Fourier series for $f(x)$.

This property of least squares may be taken as a point of departure for a discussion of the convergence of the series, under suitable hypotheses with regard to $f(x)$. While the hypotheses are more restrictive than those used in the convergence proofs already given, the method is of interest for its own sake, and furthermore lends itself to a remarkable variety of generalizations in other directions. It depends on theorems of approximation from Chapters I and II, and also on a proposition, due in the first instance to S. Bernstein, the simplicity and importance of which are noteworthy in the highest degree:

BERNSTEIN'S THEOREM. *If $T_n(x)$ is a trigonometric sum of the nth order, and if L is the maximum of its absolute value, the absolute value of the derivative $T_n'(x)$ can not exceed nL.*

The proof to be given here was devised independently by Marcel Riesz and by de la Vallée Poussin. For the purposes

of the demonstration, the theorem may be restated in the following equivalent form:

If $T_n(x)$ is a trigonometric sum of the nth order, and if the maximum of $|T_n'(x)|$ is 1, the maximum of $|T_n(x)|$ itself can not be less than $1/n$.

Suppose the maximum of $|T_n(x)|$ were less than $1/n$. Then, for any value of the constant C, the function

$$R_n(x) = \frac{1}{n}\cos(nx - C) - T_n(x)$$

would have the same sign as $\cos(nx-C)$ at each of the points C/n, $(C+\pi)/n$, $(C+2\pi)/n$, $\cdots$, $(C/n)+2\pi$, where $\cos(nx-C)$ takes on the values $+1$ and -1 alternately. Hence $R_n(x)$ would vanish at least once in each of the $2n$ intervals between successive points of this set, say at the points $x_1, x_2, \cdots, x_{2n}$, all contained in an interval of length 2π. By Rolle's theorem, $R_n'(x)$ would vanish between x_1 and x_2, between x_2 and x_3, etc., and also, as $R_n(x_1+2\pi) = R_n(x_1) = 0$, between x_{2n} and $x_1+2\pi$; and the $2n$ distinct roots of R_n' thus specified all lie within the interval from x_1 to $x_1+2\pi$. Explicitly,

$$R_n'(x) = -\sin(nx - C) - T_n'(x).$$

If C is chosen so as to make $\sin(nx-C)$ equal to the negative of $T_n'(x)$ at a point where $T_n'(x) = \pm 1$, $R_n'(x)$ will have a double root at this point; for each of the functions $\sin(nx-C)$, $T_n'(x)$ separately attains a maximum or a minimum there. Having $2n$ distinct roots in a period, and at least one double root, R_n' has roots of aggregate multiplicity at least $2n+1$. But this would require that R_n', as a trigonometric sum of order n, vanish identically, which is impossible, since R_n is sometimes positive and sometimes negative, by hypothesis, and so can not be constant. Consequently the assumption that $|T_n(x)|$ remains less than $1/n$ is inadmissible.

(The fact that a trigonometric sum of order n can not have roots of multiplicities aggregating more than $2n$ without vanishing identically is assumed here as known. A special

case of it, sufficient for the present application, is proved incidentally in another connection in Chapter IV below.)

To proceed with the proof of convergence, let $f(x)$ be a continuous function of period 2π, and $S_n(x)$ the partial sum of its Fourier series to terms of order n, and let

$$\gamma_n = \int_{-\pi}^{\pi} [f(x) - S_n(x)]^2 \, dx.$$

Let $t_n(x)$ be a trigonometric sum of the nth order, subsequently to be chosen so as to furnish an approximation to $f(x)$, but arbitrary for the present, and let

$$f(x) - t_n(x) = r(x), \qquad S_n(x) - t_n(x) = \tau_n(x).$$

Then $r - \tau_n$ is identical with $f - S_n$, and

$$\int_{-\pi}^{\pi} [r(x) - \tau_n(x)]^2 \, dx = \gamma_n.$$

Let ε_n be an upper bound for the absolute value of $r(x)$:

$$r(x) = |f(x) - t_n(x)| \leqq \varepsilon_n$$

for all values of x. Finally, let μ_n be the maximum of $|\tau_n(x)|$, and x_0 a point where $|\tau_n(x_0)| = \mu_n$.

By Bernstein's theorem.

$$|\tau_n'(x)| \leqq n\mu_n$$

for all values of x. For $|x - x_0| \leqq 1/(2n)$, consequently,

$$|\tau_n(x) - \tau_n(x_0)| \leqq \frac{1}{2}\mu_n, \qquad |\tau_n(x)| \geqq \frac{1}{2}\mu_n.$$

If $\varepsilon_n \leqq \frac{1}{4}\mu_n$ (the contrary case being reserved for separate mention),

$$|r(x)| \leqq \frac{1}{4}\mu_n, \qquad |r(x) - \tau_n(x)| \geqq \frac{1}{4}\mu_n,$$

throughout the interval specified. As the length of the interval is $1/n$,

$$\gamma_n = \int_{-\pi}^{\pi} [r(x) - \tau_n(x)]^2 \, dx > \frac{1}{n} \left(\frac{\mu_n}{4}\right)^2,$$

whence $\mu_n \leqq 4(n\gamma_n)^{1/2}$. To suppose, on the other hand, that $\epsilon_n > \frac{1}{4}\mu_n$, is to assume outright that $\mu_n < 4\epsilon_n$. So in either case μ_n does not exceed the larger of the quantities $4\epsilon_n$, $4(n\gamma_n)^{1/2}$, or, in a single formula,

$$\mu_n \leqq 4(n\gamma_n)^{1/2} + 4\epsilon_n.$$

But $|\tau_n| \leqq \mu_n$, $|r| \leqq \epsilon_n$, and $r - \tau_n$ is the same as $f - S_n$. Consequently

$$|f(x) - S_n(x)| \leqq 4(n\gamma_n)^{1/2} + 5\epsilon_n$$

for all values of x.

The function $f(x)$ being continuous, it is always possible to choose $t_n(x)$ for successive values of n in such a way that $\lim_{n=\infty} \epsilon_n = 0$. So the relation just obtained amounts to a proof of

THEOREM II. *If $f(x)$ is a continuous function of period 2π, its Fourier series will converge uniformly to the value $f(x)$, provided that*

$$\lim_{n=\infty} n\gamma_n = 0.$$

In particular, Theorem VI of Chapter II yields at once

COROLLARY I. *A sufficient condition for uniform convergence is that $f(x)$ be absolutely continuous.*

By Theorem II of Chapter I, on the other hand, if $\omega(\delta)$ is the modulus of continuity of $f(x)$, there will exist sums $t_n(x)$ such that

$$|f(x) - t_n(x)| \leqq K'\omega(2\pi/n),$$

where K' is an absolute constant. Then

$$\int_{-\pi}^{\pi} [f(x) - t_n(x)]^2 \, dx \leqq 2\pi [K'\omega(2\pi/n)]^2.$$

But as the integral takes on its minimum value when $S_n(x)$ is put in place of $t_n(x)$, the right-hand member of the last relation is an upper bound for γ_n. So $n\gamma_n$ will approach zero if $n^{1/2}\omega(2\pi/n)$ approaches zero, and it is possible to state

COROLLARY II. *A sufficient condition for uniform convergence is that* $\lim_{\delta=0} \omega(\delta)/\delta^{1/2} = 0$, *where* $\omega(\delta)$ *is the modulus of continuity of* $f(x)$.

The conditions of Corollaries I and II of course overlap to a considerable extent, but they are not coextensive, and neither includes the other. As previously mentioned, they are less general than others which have already been obtained; absolute continuity is more stringent than limited variation, and the requirement involving $\delta^{1/2}$ is more restrictive than the Lipschitz-Dini condition.

The novel possibilities of the present method will be more apparent, if it is remarked that in the proof of Theorem II, down through the final relation of inequality preceding the statement of the theorem, no use whatever was made of the assumption that $S_n(x)$ was the Fourier sum for $f(x)$, and the value of the integral γ_n a minimum; the argument can be repeated step by step with $S_n(x)$ replaced by $T_n(x)$, an arbitrary trigonometric sum of the nth order, and γ_n replaced by g_n, the integral of $[f(x) - T_n(x)]^2$ over a period. The conclusion is

THEOREM IIa. *If* $f(x)$ *is a continuous function of period* 2π, $T_n(x)$ *an arbitrary trigonometric sum of the nth order, and*

$$g_n = \int_{-\pi}^{\pi} [f(x) - T_n(x)]^2 \, dx,$$

and if there exists a trigonometric sum $t_n(x)$, *of the nth order, such that*

$$|f(x) - t_n(x)| \leq \epsilon_n$$

for all values of x, *then, for all values of* x,

$$|f(x) - T_n(x)| \leq 4(n g_n)^{1/2} + 5\epsilon_n.$$

To this may be added immediately

THEOREM IIb. *If sums* $T_n(x)$ *are defined for infinitely many values of* n *in such a way that* $g_n \leq A\gamma_n$, *where* A *is independent of* n, *the sums* $T_n(x)$ *will converge uniformly to*

the value $f(x)$, as n becomes infinite, under the conditions of Theorem II *or either of its corollaries.*

The next application depends for its simplest expression on an existence theorem, the proof of which will be postponed for a few pages, to avoid interruption of the current order of ideas. The statement of the theorem is as follows:

THEOREM III. *If $\varrho(x)$ is a summable function of period 2π which is nowhere negative, and is different from zero over a point set of positive measure in a period, if $f(x)$ is a summable function of period 2π such that $\varrho(x)[f(x)]^2$ is also summable, and if $T_n(x)$ is a trigonometric sum of specified order n, the integral*

$$\int_{-\pi}^{\pi} \varrho(x)\,[f(x) - T_n(x)]^2\,dx$$

has a minimum, which is attained for one and just one determination of the coefficients in $T_n(x)$.

The truth of this assertion being assumed for the moment, attention will be directed to the question of the convergence of $T_n(x)$ toward $f(x)$ as n becomes infinite, the function $f(x)$ being supposed continuous. The problem is connected with an extensive theory developed from a different point of view by Stekloff, Szegö, and others. The minimizing sum may be called the *approximating sum of order n corresponding to the weight function $\varrho(x)$*, and it will be understood now that $T_n(x)$ denotes this particular sum. The corresponding minimum value of the integral will be denoted by G_n. For the convergence proof, let $\varrho(x)$ be further restricted by the hypothesis that its values are comprised between two positive bounds:

$$0 < v \leqq \varrho(x) \leqq V,$$

where v and V are independent of x.

By the minimizing property of $T_n(x)$,

$$\begin{aligned} G_n &= \int_{-\pi}^{\pi} \varrho(x)\,[f(x) - T_n(x)]^2\,dx \\ &\leqq \int_{-\pi}^{\pi} \varrho(x)\,[f(x) - S_n(x)]^2\,dx \leqq V\gamma_n, \end{aligned}$$

where $S_n(x)$ and γ_n have their previous meanings. On the other hand, with the notation of Theorem IIa,

$$G_n \geqq v\int_{-\pi}^{\pi}[f(x) - T_n(x)]^2\,dx = vg_n.$$

So $g_n \leqq (V/v)\gamma_n$, and the conclusions of Theorem IIb are applicable:

THEOREM IV. *If the weight function $\varrho(x)$ has a finite upper bound and a positive lower bound, the approximating sum $T_n(x)$ will converge uniformly toward $f(x)$ as n becomes infinite, if* $\lim_{n=\infty} n\gamma_n = 0$, *and hence, in particular, if $f(x)$ is absolutely continuous, or has a modulus of continuity $\omega(\delta)$ such that* $\lim_{\delta=0}\omega(\delta)/\delta^{1/2} = 0$.

In modification of this result, let the restriction $|\varrho(x)| \leqq V$ be removed, the hypothesis with regard to ϱ being merely that it is summable, and that $\varrho \geqq v > 0$ everywhere. Since G_n is a minimum,

$$G_n \leqq \int_{-\pi}^{\pi}\varrho(x)\,[f(x) - t_n(x)]^2\,dx$$

if $t_n(x)$ is *any* trigonometric sum of the nth order. Let ε_n be an upper bound of $|f - t_n|$, as before. Then

$$G_n \leqq \varepsilon_n^2\int_{-\pi}^{\pi}\varrho(x)\,dx = I\varepsilon_n^2,$$

with the notation $\int\varrho = I$. As it is still true that $G_n \geqq vg_n$, it follows that $g_n \leqq (I/v)\,\varepsilon_n^2$. Under the hypothesis that $\lim_{\delta=0}\omega(\delta)/\delta^{1/2} = 0$, $t_n(x)$ can be chosen so that $\lim_{n=\infty} n^{1/2}\varepsilon_n = 0$, which means that $\lim_{n=\infty} ng_n = 0$, and Theorem IIa shows that $T_n(x)$ *converges uniformly to $f(x)$ once more.*

2. Convergence of trigonometric approximation as related to integral of *m*th power of error

To generalize in another direction, consider the integral

$$\int_{-\pi}^{\pi}\varrho(x)\,|f(x) - T_n(x)|^m\,dx,$$

in which m is an arbitrary positive number, taking the place of the particular exponent $m = 2$ that has been used hitherto.

When $m, n, \varrho(x)$, and $f(x)$ are given, $f(x)$ being bounded and measurable, it can be shown that there exists a determination of the coefficients in $T_n(x)$ for which the value of the integral is a minimum, provided that $\varrho(x)$ is summable, everywhere positive or zero, and different from zero over a point set of positive measure in a period. When $m > 1$, the determination is unique. The existing presentations do not give the theorem with quite this degree of generality, but it is not difficult to supply the necessary extensions. For somewhat less general functions ϱ (in particular, for the case that $\varrho = 1$ identically) it has been shown that the determination is unique for $m = 1$ also, if $f(x)$ is continuous. It is not generally unique when $m < 1$. The proofs will be omitted here, the facts with regard to convergence being formulated in such a way as not to presuppose a knowledge of the theorems of existence and uniqueness.

For the discussion of convergence, let $f(x)$ be a continuous function of period 2π, and let it be supposed that there is a number v such that

$$\varrho(x) \geqq v > 0$$

for all values of x, while $\varrho(x)$, as always, is summable over a period. Let

$$g_n = \int_{-\pi}^{\pi} \varrho(x)\,|f(x) - T_n(x)|^m\,dx,$$

$T_n(x)$ being for the present an arbitrary trigonometric sum of the nth order. (The notation is changed somewhat from that previously used in treating the case $m = 2$.) Let it be assumed that there is a sum $t_n(x)$, likewise of the nth order, such that

$$|f(x) - t_n(x)| \leqq \varepsilon_n$$

for all values of x, and let

$$f(x) - t_n(x) = r(x), \qquad T_n(x) - t_n(x) = \tau_n(x),$$

whence

$$\int_{-\pi}^{\pi} \varrho(x)\,|r(x) - \tau_n(x)|^m\,dx = g_n.$$

If $\mu_n = |\tau_n(x_0)|$ is the maximum of $|\tau_n(x)|$, it follows that $|\tau_n'(x)| \leq n\mu_n$, by Bernstein's theorem once more, and there is an interval of length at least $1/n$ throughout which

$$|\tau_n(x) - \tau_n(x_0)| \leq \frac{1}{2}\mu_n, \qquad |\tau_n(x)| \geq \frac{1}{2}\mu_n,$$

$$|r(x) - \tau_n(x)| \geq \frac{1}{4}\mu_n,$$

provided that $\varepsilon_n \leq \frac{1}{4}\mu_n$. Then

$$g_n \geq \frac{v}{n}\left(\frac{\mu_n}{4}\right)^m, \qquad \mu_n \leq 4\,(n g_n/v)^{1/m}.$$

Whether $\varepsilon_n \leq \frac{1}{4}\mu_n$ or not,

$$\mu_n \leq 4\,(n g_n/v)^{1/m} + 4\,\varepsilon_n,$$

and

$$|f(x) - T_n(x)| \leq |r(x) - \tau_n(x)| \leq 4\,(n\,g_n/v)^{1/m} + 5\,\varepsilon_n.$$

For each value of n, let γ_n be the greatest lower bound of the integral g_n, when all possible values are given to the coefficients in $T_n(x)$, while the functions ϱ and f and the exponent m are held fast. It is clear that $\gamma_n \geq 0$. Nothing is assumed as to the possibility of making g_n actually equal to γ_n. by one or more determinations of the coefficients. As g_n can not be less than γ_n for any sum of the nth order, it follows in particular, since $t_n(x)$ is such a sum, that

$$\gamma_n \leq \int_{-\pi}^{\pi} \varrho(x)\,|f(x) - t_n(x)|^m\,dx \leq I\,\varepsilon_n^m,$$

the value of $\int \varrho$ being denoted by I. If sums $T_n(x)$ are chosen for the successive values of n so that $g_n \leq A\gamma_n$, where A is independent of n, then $g_n^{1/m} \leq (AI)^{1/m}\,\varepsilon_n$, and

$$|f(x) - T_n(x)| \leq B n^{1/m}\varepsilon_n,$$

where B is likewise independent of n. The sums $T_n(x)$ will converge uniformly toward $f(x)$, if $n^{1/m}\varepsilon_n$ can be made to approach zero, and this will be possible under conditions indicated by Theorem II of Chapter I, if $m > 1$; by the Corollary of Theorem IV of that chapter, if $m = 1$, and by

Theorem IV itself, if $0 < m < 1$. The principal results may be summarized in the following formal statement:

THEOREM V. *If the weight function $\varrho(x)$ is summable and has a positive lower bound, and if a sum $T_n(x)$ of the nth order is chosen for each of an infinite set of values of n so that $g_n \leq A\gamma_n$, where A is independent of n, these sums will converge uniformly toward $f(x)$, when m is held fast and n is allowed to become infinite, if $m > 1$ and $f(x)$ has a modulus of continuity $\omega(\delta)$ such that $\lim_{\delta \doteq 0} \omega(\delta)/\delta^{1/m} = 0$, or if $m = 1$ and $f(x)$ has a continuous derivative.*

The explicit statement for $m < 1$ is more complicated and less interesting. In the cases covered by the theorem as stated, for $m \geq 1$, there is in fact a determinate approximating sum for each value of n, making $g_n = \gamma_n$, and the conclusions of course apply in particular to the convergence of these approximating sums. On the other hand, it is clear that A can be replaced by a factor A_n which becomes infinite with n, if more restrictive hypotheses are placed on $f(x)$, so that ϵ_n can be made to approach zero with sufficiently increased rapidity.

3. Proof of an existence theorem

To carry out an intention expressed above, it remains to supply a proof of Theorem III, with regard to the existence of an approximating sum corresponding to a given weight function $\varrho(x)$ when $m = 2$.

It will be worth while to begin with more general considerations. Let $q_0(x)$, $q_1(x)$, $q_2(x)$, $\cdots$, be a sequence of functions, each summable together with its square over an interval (a, b). Let these functions be linearly independent, in the sense that any linear combination of a finite number of them, with coefficients not all zero, is different from zero *over a point set of positive measure* in (a, b), which is equivalent to saying that the integral of the square of such a linear combination, extended from a to b, is positive. The requirement may be indicated briefly by saying that the q's are *properly independent.* Let

$$Q_0 = q_0, \qquad Q_1 = q_1 - Q_0 \frac{\int q_1 Q_0}{\int Q_0^2},$$

$$Q_2 = q_2 - Q_0 \frac{\int q_2 Q_0}{\int Q_0^2} - Q_1 \frac{\int q_2 Q_1}{\int Q_1^2},$$

and generally

$$Q_k = q_k - \sum_{j=0}^{k-1} Q_j \frac{\int q_k Q_j}{\int Q_j^2};$$

it is understood that each integration is to be performed with regard to x, from a to b. Each denominator is different from zero, because of the hypothesis of proper independence. It is seen from the definition of Q_1 that $\int Q_0 Q_1 = 0$. Since this is the case, the definition of Q_2 makes $\int Q_0 Q_2 = \int Q_1 Q_2 = 0$. Generally, each Q is orthogonal to all the preceding Q's, or, in other words, any two of the Q's are orthogonal to each other. If

$$p_k(x) = Q_k \Big/ \left(\int Q_k^2\right)^{1/2},$$

the functions p_k form a normalized orthogonal sequence, to which Theorem I is applicable. Each p_k is a linear combination of $q_0, \cdots, q_k$, and conversely each q_k can be linearly expressed in terms of p_k and the p's of lower order, since the coefficient of q_k in the expression for p_k is not zero.

Now let $\varrho(x)$ be a non-negative summable function of period 2π, such that

$$\int_{-\pi}^{\pi} \varrho(x)\, dx > 0,$$

and let $[\varrho(x)]^{1/2} = w(x)$. The functions

$$w(x), \quad w(x)\cos x, \quad w(x)\sin x, \quad w(x)\cos 2x, \quad w(x)\sin 2x, \cdots$$

are summable, with their squares, and properly independent; a linear combination of a finite number of them, with co-

efficients not all zero, is a trigonometric sum multiplied by $w(x)$, and is different from zero wherever $\varrho \neq 0$, with the exception of a finite number of points at most in a period. If they are taken as the functions $q_0, q_1, q_2, \cdots$, of the preceding paragraph, over the interval $(-\pi, \pi)$, the corresponding functions $p_0, p_1, p_2, \cdots$. have the form

$$w(x)\,U_0(x), \quad w(x)\,U_1(x), \quad w(x)\,V_1(x), \quad w(x)\,U_2(x),$$
$$w(x)\,V_2(x), \cdots,$$

where U_k and V_k are trigonometric sums of the kth order, and

$$\int_{-\pi}^{\pi} \varrho(x)\,U_j(x)\,U_k(x)\,dx = \int_{-\pi}^{\pi} \varrho(x)\,V_j(x)\,V_k(x)\,dx = 0 \quad (j \neq k),$$

$$\int_{-\pi}^{\pi} \varrho(x)\,U_j(x)\,V_k(x)\,dx = 0 \qquad \text{for all } j \text{ and } k,$$

$$\int_{-\pi}^{\pi} \varrho\,(x)\,[U_k(x)]^2\,dx = \int_{-\pi}^{\pi} \varrho(x)\,[V_k(x)]^2\,dx = 1.$$

If $T_n(x)$ is any trigonometric sum of the nth order, $w(x)\,T_n(x)$ is a linear combination of the functions now taken as the q's, and so is a linear combination of the functions $w(x)\,U_k(x)$, $w(x)\,V_k(x)$, $k = 0, 1, \cdots, n$, and T_n itself is expressible in terms of U_k, V_k, with the same coefficients.

Let $f(x)$ be a summable function of period 2π, such that $\varrho(x)\,[f(x)]^2$ is also summable. Then

$$\int_{-\pi}^{\pi} \varrho(x)\,[f(x) - T_n(x)]^2\,dx = \int_{-\pi}^{\pi} [w(x)\,f(x) - w(x)\,T_n(x)]^2\,dx,$$

and Theorem I can be applied, with $w(x)\,f(x)$ as the function for which an approximation is sought. The integral has a minimum value, and if

$$T_n(x) = A_0\,U_0(x) + A_1\,U_1(x) + \cdots + A_n\,U_n(x)$$
$$+ B_1\,V_1(x) + \cdots + B_n\,V_n(x),$$

it is necessary and sufficient for the attainment of the minimum that

$$A_k = \int_{-\pi}^{\pi} w(x) f(x) \cdot w(x)\, U_k(x)\, dx = \int_{-\pi}^{\pi} \varrho(x) f(x)\, U_k(x)\, dx,$$
$$B_k = \int_{-\pi}^{\pi} \varrho(x) f(x)\, V_k(x)\, dx.$$

4. Polynomial approximation

Attention will next be directed to problems of polynomial approximation. An essential instrument here is the form of Bernstein's theorem in which Bernstein himself was primarily interested:

Bernstein's Theorem for Polynomials. *If $P_n(x)$ is a polynomial of the nth degree, and L the maximum of $|P_n(x)|$ for $-1 \leqq x \leqq 1$, then*

$$|P_n'(x)| \leqq \frac{nL}{(1-x^2)^{1/2}}$$

for $-1 < x < 1$.

It can be deduced immediately from the theorem for the trigonometric case. Let $x = \cos\theta$. Then $P_n(x) = P_n(\cos\theta)$ is a trigonometric sum of the nth order in θ, having L for the maximum of its absolute value, and consequently

$$\left|\frac{d}{d\theta} P_n(\cos\theta)\right| = |\sin\theta\, P_n'(\cos\theta)| = |(1-x^2)^{1/2} P_n'(x)| \leqq nL,$$

whence $|P_n'(x)| \leqq nL/(1-x^2)^{1/2}$.

If L is the maximum of $|P_n(x)|$ for $a \leqq x \leqq b$, let

$$y = \frac{2x-a-b}{b-a}.$$

Then $P_n(x) = Q_n(y)$ is a polynomial of the nth degree in y, having L for the maximum of its absolute value in the interval $-1 \leqq y \leqq 1$. Therefore

$$\left|\frac{d}{dy} Q_n(y)\right| = \left|P_n'(x) \cdot \frac{dx}{dy}\right| = \frac{b-a}{2} |P_n'(x)| < \frac{nL}{(1-y^2)^{1/2}};$$

by substitution of the value of y in terms of x in the factors $(1-y)$, $(1+y)$ under the radical sign, one is led to

Corollary I. *If $P_n(x)$ is a polynomial of the nth degree, and L the maximum of $|P_n(x)|$ for $a \leqq x \leqq b$, then*

$$|P_n'(x)| \leq \frac{nL}{[(b-x)(x-a)]^{1/2}}$$

for $a < x < b$.

This implies directly

Corollary II. *If* $P_n(x)$ *is a polynomial of the nth degree, and* L *the maximum of* $|P_n(x)|$ *for* $a \leq x \leq b$, *and if* $a < a_1 < b_1 < b$, *then*

$$|P_n'(x)| \leq hnL$$

for $a_1 \leq x \leq b_1$, *where* h *depends only on* a, a_1, b_1, *and* b.

A further corollary is to be obtained by the following considerations. Let $P_n(x)$ once more be a polynomial of the nth degree, and L the maximum of its absolute value for $-1 \leq x < 1$. Let x be a number of the interval $-1 < x \leq 1$, and δ a positive number ≤ 1 such that $-1 \leq x-\delta < 1$. Then

$$|P_n(x) - P_n(x-\delta)| = \left|\int_{x-\delta}^{x} P_n'(t)\,dt\right| \leq nL\int_{x-\delta}^{x} \frac{dt}{(1-t^2)^{1/2}}.$$

It is readily seen that

$$\int_{x-\delta}^{x} \frac{dt}{(1-t^2)^{1/2}} \leq \int_{1-\delta}^{1} \frac{dt}{(1-t^2)^{1/2}}.$$

(For an analytical proof, let $\delta = 2\eta$, $x - \frac{1}{2}\delta = y$, so that $x-\delta = y-\eta$, $x = y+\eta$,

$$\int_{x-\delta}^{x} \frac{dt}{(1-t^2)^{1/2}} = \int_{y-\eta}^{y+\eta} \frac{dt}{(1-t^2)^{1/2}}.$$

Since

$$\frac{d}{dy}\int_{y-\eta}^{y+\eta} \frac{dt}{(1-t^2)^{1/2}} = \frac{1}{[1-(y+\eta)^2]^{1/2}} - \frac{1}{[1-(y-\eta)^2]^{1/2}},$$

which is equal to 0 if $y = 0$, positive if $y > 0$, negative if $y < 0$, the value of the integral is a minimum when $y = 0$, that is, when the interval of integration is symmetric about the origin, and increases steadily as the interval is displaced toward either side.) As $\delta \leq 1$, $0 \leq t < 1$ in the integral from $1-\delta$ to 1, $1-t^2 \geq 1-t \geq 0$, and

$$\int_{1-\delta}^{1} \frac{dt}{(1-t^2)^{1/2}} \leqq \int_{1-\delta}^{1} \frac{dt}{(1-t)^{1/2}} = 2\delta^{1/2}.$$

So

$$|P_n(x) - P_n(x-\delta)| \leqq 2nL\delta^{1/2}.$$

An equivalent statement is that if x_1 and x_2 are two numbers in $(-1, 1)$, differing by not more than unity,

$$|P_n(x_2) - P_n(x_1)| \leqq 2nL|x_2 - x_1|^{1/2}.$$

Now suppose that L is the maximum of $|P_n(x)|$ for $a \leqq x \leqq b$, and let x_1 and x_2 be two numbers of this interval, differing by not more than $\frac{1}{2}(b-a)$. Let $y = (2x-a-b)/(b-a)$, the values of y corresponding to x_1 and x_2 being y_1 and y_2, and let $P_n(x) = Q_n(y)$. Then L is the maximum of $|Q_n(y)|$ for $-1 \leqq y \leqq 1$, y_1 and y_2 are two numbers of this interval, differing by not more than unity, and

$$\begin{aligned} |P_n(x_2) - P_n(x_1)| &= |Q_n(y_2) - Q_n(y_1)| \leqq 2nL|y_2 - y_1|^{1/2} \\ &= 2nL[2|x_2 - x_1|/(b-a)]^{1/2}. \end{aligned}$$

The conclusion is

COROLLARY III. *If $P_n(x)$ is a polynomial of the nth degree, and L the maximum of $|P_n(x)|$ for $a \leqq x \leqq b$, and if x_1 and x_2 are two numbers in the interval, differing by not more than $\frac{1}{2}(b-a)$, then*

$$|P_n(x_2) - P_n(x_1)| \leqq HnL|x_2 - x_1|^{1/2},$$

where $H = 2/[\frac{1}{2}(b-a)]^{1/2}$.

By the introduction of an intermediate value $x_3 = \frac{1}{2}(x_1 + x_2)$, it may be seen that

$$|P_n(x_2) - P_n(x_1)| \leqq \frac{4nL}{(b-a)^{1/2}}|x_2 - x_1|^{1/2},$$

without the restriction that $|x_2 - x_1| \leqq \frac{1}{2}(b-a)$.

The way has now been prepared for a discussion of problems of approximation associated with the integral

$$\int_a^b \varrho(x)|f(x) - P_n(x)|^m dx,$$

where $\varrho(x)$ and $f(x)$ are given functions in the interval (a, b), $P_n(x)$ is a polynomial of the nth degree, and m is a given exponent >0.

When $m = 2$, a theorem of existence and uniqueness may be stated as follows:

THEOREM VI. *If $\varrho(x)$ is a summable function which is nowhere negative, and is different from zero over a point set of positive measure in the interval (a, b), if $f(x)$ is a summable function such that $\varrho(x)[f(x)]^2$ is also summable from a to b, and if $P_n(x)$ is a polynomial of specified degree n, the integral*

$$\int_a^b \varrho(x)[f(x) - P_n(x)]^2 dx$$

has a minimum, which is attained for one and just one determination of the coefficients in $P_n(x)$.

The proof is entirely analogous to that of Theorem III. The functions

$$w(x),\ x w(x),\ x^2 w(x),\ \cdots,$$

where $w(x) = [\varrho(x)]^{1/2}$, are summable together with their squares and properly independent over the interval (a, b). If they are taken as the functions $q_k(x)$ previously considered. the corresponding functions $p_k(x)$ have the form $w(x)\,Y_k(x)$. where $Y_k(x)$ is a polynomial of the kth degree, and

$$\int_a^b \varrho(x)\,Y_j(x)\,Y_k(x)\,dx = 0 \qquad (j \neq k),$$

$$\int_a^b \varrho(x)\,[Y_k(x)]^2 dx = 1.$$

The polynomials $Y_k(x)$ are called the *Tchebychef polynomials corresponding to the characteristic function $\varrho(x)$ in the interval (a, b)*. Any polynomial $P_n(x)$, of the nth degree, can be expressed as a linear combination of $Y_0, Y_1, \cdots, Y_n$, and for the minimizing of the integral in the statement of the theorem it is necessary and sufficient that the coefficient of Y_k in this expression be

$$\int_a^b \varrho(x)\,f(x)\,Y_k(x)\,dx.$$

In the case that ϱ is identically equal to 1 and the interval is that from -1 to $+1$, $Y_k(x)$ is equal to $(k+\frac{1}{2})^{1/2} X_k(x)$, where X_k is the Legendre polynomial of the kth degree; and the approximating polynomial $P_n(x)$ then is the partial sum of the Legendre series for $f(x)$.

For an arbitrary value of $m > 0$, theorems of existence and uniqueness can again be established in analogy with those of the trigonometric case. An approximating polynomial, reducing the integral to a minimum, exists if $f(x)$ is bounded and measurable, and if $\varrho(x)$ is summable, nowhere negative, and positive over a point set of positive measure in (a, b). The approximating polynomial is unique when $m > 1$, and, with suitably restricted hypotheses, when $m = 1$ also. The proofs will not be given here, the further discussion being put in such a form that a knowledge of them is not essential.

Let $f(x)$ be a continuous function for $a \leqq x \leqq b$, and $P_n(x)$ an arbitrary polynomial of the nth degree. Let $\varrho(x)$ be summable over (a, b), and nowhere negative in the interval, and let $\varrho(x) \geqq v > 0$ for $\alpha_0 \leqq x \leqq \beta_0$, where $a \leqq \alpha_0 < \beta_0 \leqq b$. (It will be noticed, as a departure from the conditions of the earlier convergence proofs, that the hypothesis $\varrho \geqq v > 0$ does not necessarily hold over the entire range of variation of x.) Let

$$g_n = \int_a^b \varrho(x) \, |f(x) - P_n(x)|^m \, dx.$$

Furthermore, let $p_n(x)$ be a polynomial of the nth degree, in general distinct from $P_n(x)$,

$$|f(x) - p_n(x)| \leqq \epsilon_n$$

for $a \leqq x \leqq b$, and

$$f(x) - p_n(x) = r(x), \qquad P_n(x) - p_n(x) = \pi_n(x),$$

so that

$$\int_a^b \varrho(x) \, |r(x) - \pi_n(x)|^m \, dx = g_n.$$

Suppose that μ_n is the maximum of $|\pi_n(x)|$ for $\alpha_0 \leqq x \leqq \beta_0$, and that x_0 is a point in (α_0, β_0) at which this maximum is attained. By Corollary III above,

$$|\pi_n(x) - \pi_n(x_0)| \leqq H n \mu_n |x - x_0|^{1/2}$$

as long as x is in (α_0, β_0) and distant from x_0 by not more than $\frac{1}{2}(\beta_0 - \alpha_0)$, the factor H being equal to $2/[\frac{1}{2}(\beta_0 - \alpha_0)]^{1/2}$. These conditions will be satisfied for $n \geqq 1$ throughout one or the other (or both) of the intervals $(x_0 - \delta, x_0)$, $(x_0, x_0 + \delta)$, if

$$\delta = \frac{1}{4H^2 n^2} = \frac{\beta_0 - \alpha_0}{32 n^2}.$$

Over such an interval, then, since $|x - x_0|^{1/2} \leqq 1/(2Hn)$,

$$|\pi_n(x) - \pi_n(x_0)| \leqq \frac{1}{2}\mu_n, \qquad |\pi_n(x)| \geqq \frac{1}{2}\mu_n.$$

If $\epsilon_n < \frac{1}{4}\mu_n$,

$$|r(x) - \pi_n(x)| \geqq \frac{1}{4}\mu_n$$

over the same range; and, as the length of the interval is $1/(4H^2n^2)$,

$$g_n \geqq \frac{v}{4H^2n^2}\left(\frac{\mu_n}{4}\right)^m,$$

$$\mu_n < 4(4H^2/v)^{1/m}(n^2 g_n)^{1/m} = B_0(n^2 g_n)^{1/m},$$

with $B_0 = 4(4H^2/v)^{1/m}$. Whether $\epsilon_n \leqq \frac{1}{4}\mu_n$ or not,

$$\mu_n \leqq B_0(n^2 g_n)^{1/m} + 4\epsilon_n,$$

$$|f(x) - P_n(x)| = |r(x) - \pi_n(x)| \leqq B_0(n^2 g_n)^{1/m} + 5\epsilon_n$$

throughout (α_0, β_0).

Let γ_n be the greatest lower bound of g_n, when $\varrho(x)$, $f(x)$, and m are given. As an immediate consequence of the definition,

$$\gamma_n \leqq \int_a^b \varrho(x)\,|f(x) - p_n(x)|^m\,dx \leqq I\epsilon_n^m, \qquad I = \int_a^b \varrho(x)\,dx.$$

If polynomials $P_n(x)$ are chosen for an infinite succession of values of n so that $g_n \leqq A\gamma_n$, where A is independent of n, then $g_n^{1/m} \leqq (AI)^{1/m}\epsilon_n$, and

$$|f(x) - P_n(x)| \leqq B n^{2/m}\epsilon_n,$$

7

for $\alpha_0 \leqq x \leqq \beta_0$, with a factor B which is likewise independent of n.

The conditions to be imposed on $f(x)$, in order that it may be possible to make $n^{2/m}\epsilon_n$ approach zero, are to be obtained from Theorems VI and VIII of Chapter I, and the Corollary of the latter. It will be sufficient to state the results formally for $m \geqq 2$, though corresponding statements could readily be added for $0 < m < 2$:

THEOREM VII. *If $\varrho(x)$ is summable over (a, b), and has a positive lower bound for $\alpha_0 \leqq x \leqq \beta_0$, where $a \leqq \alpha_0 < \beta_0 \leqq b$, and if a polynomial $P_n(x)$ is chosen for each of an infinite set of values of n so that $g_n \leqq A\gamma_n$, where A is independent of n, these polynomials will converge uniformly toward $f(x)$ for $\alpha_0 \leqq x \leqq \beta_0$, when m is held fast and n is allowed to become infinite, if $m > 2$ and $f(x)$ has a modulus of continuity $\omega(\delta)$ such that $\lim_{\delta=0}\omega(\delta)/\delta^{2/m} = 0$, or if $m = 2$ and $f(x)$ has a continuous derivative.*

Under the conditions formulated, there will in fact be one and just one polynomial of each degree n for which $g_n = \gamma_n$, not merely for $m = 2$, in accordance with Theorem VI, but also for an arbitrary $m > 2$, under a theorem cited above without proof. Theorem VII naturally applies to the convergence of these approximating polynomials. More particularly still, for $\varrho(x) = 1$, $a = -1$, $b = 1$, one may state

COROLLARY I. *If $f(x)$ has a continuous derivative for $-1 \leqq x \leqq 1$, its Legendre series converges uniformly toward $f(x)$ throughout the closed interval.*

Another part of the content of the theorem, aside from the question of uniformity of convergence, is expressed in

COROLLARY II. *If $\varrho(x)$ is continuous and nowhere negative for $a \leqq x \leqq b$, the hypotheses of Theorem VII remaining otherwise unchanged, the polynomials $P_n(x)$ will converge toward $f(x)$ at all points where $\varrho(x) \neq 0$.*

Additional information can be gained by further elaboration of the proof of the theorem. Let the previous notation be continued in force, with the understanding now that polynomials $P_n(x)$ and $p_n(x)$ are defined for all positive integral

values of n, and that these sequences of polynomials are kept unchanged throughout the course of the reasoning. Let α and β be two numbers such that $\alpha_0 < \alpha < \beta < \beta_0$. Let $\alpha_1, \alpha_2, \cdots$ and $\beta_1, \beta_2, \cdots$ be two infinite successions of numbers such that

$$\alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha, \qquad \beta < \cdots < \beta_2 < \beta_1 < \beta_0.$$

Let ν_n be the maximum of $|\pi_n(x)|$ for $\alpha \leq x \leq \beta$, and μ_{nk} the maximum of the same function for $\alpha_k \leq x \leq \beta_k$; μ_{n0} is the same as the previous μ_n. It follows immediately from the definitions that

$$\mu_{n0} \geq \mu_{n1} \geq \mu_{n2} \geq \cdots \geq \nu_n.$$

Until the contrary case is explicitly mentioned, it will be assumed that $\varepsilon_n \leq \frac{1}{4}\nu_n$. Then, *a fortiori*, $\varepsilon_n \leq \frac{1}{4}\mu_{n0}$, and $\mu_{n0} \leq B_0(n^2 g_n)^{1/m}$, by the earlier work. If $g_n \leq A\gamma_n$, as will be assumed from now on, $g_n^{1/m}$ does not exceed a constant multiple of ε_n, and

$$\mu_{n0} \leq A_0 n^{2/m} \varepsilon_n,$$

where A_0 is independent of n.

Suppose it is known, for a specified value of k, that

$$\mu_{nk} \leq A_k n^{\sigma} \varepsilon_n.$$

A_k being independent of n. The exponent σ will be denoted by σ_k when there is occasion to emphasize its dependence on k. By Corollary II of Bernstein's theorem,

$$|\pi_n'(x)| \leq h n \mu_{nk}$$

for $\alpha_{k+1} \leq x \leq \beta_{k+1}$, the factor h depending on α_k, β_k, α_{k+1}, and β_{k+1}, but not on anything else. Let x_1 be a point in $(\alpha_{k+1}, \beta_{k+1})$ such that $|\pi_n(x_1)| = \mu_{n,k+1}$, and let $\delta_1 = \mu_{n,k+1}/(2hn\mu_{nk})$. Because of the fact that $\mu_{n,k+1} \leq \mu_{nk}$, $\delta_1 \leq 1/(2hn)$, and it is certain that at least one of the intervals $(x_1 - \delta_1, x_1)$, $(x_1, x_1 + \delta_1)$ consists entirely of points belonging to $(\alpha_{k+1}, \beta_{k+1})$, as soon as n is sufficiently large. This condition being satisfied.

$$|\pi_n(x) - \pi_n(x_1)| \leqq \frac{1}{2}\mu_{n,k+1}, \qquad |\pi_n(x)| \geqq \frac{1}{2}\mu_{n,k+1},$$

throughout the interval designated, and as it has been supposed that $\epsilon_n \leqq \frac{1}{4}\nu_n \leqq \frac{1}{4}\mu_{n,k+1}$,

$$|r(x) - \pi_n(x)| \geqq \frac{1}{4}\mu_{n,k+1}, \quad g_n \geqq v\delta_1\left(\frac{\mu_{n,k+1}}{4}\right)^m = \frac{v\,\mu_{n,k+1}^{m+1}}{4^m \cdot 2hn\mu_{nk}},$$

$$\mu_{n,k+1} \leqq B_{k+1}(n g_n \mu_{nk})^{1/(m+1)},$$

where B_{k+1} is independent of n. Taken with the relations $g_n \leqq AI\epsilon_n^m$, $\mu_{nk} \leqq A_k n^\sigma \epsilon_n$, this means that

$$\mu_{n,k+1} \leqq B_{k+1}(A A_k I n^{\sigma+1} \epsilon_n^{m+1})^{1/(m+1)},$$

or, if $B_{k+1}(A A_k I)^{1/(m+1)}$ is denoted by A_{k+1},

$$\mu_{n,k+1} \leqq A_{k+1} n^{(\sigma+1)/(m+1)} \epsilon_n.$$

Because of a qualification introduced at one stage in the work, the result has been obtained in the first instance only for values of n from a certain point on; but it can be extended to the finite number of values of n previously neglected, from $n = 1$ on, by a suitable adjustment of the value of A_{k+1}. It is recognized therefore that the successive numbers μ_{nk} have upper bounds involving a sequence of exponents σ_k, beginning with $\sigma_0 = 2/m$, and so related that $\sigma_{k+1} = (\sigma_k + 1)/(m+1)$.

It may readily be verified by induction that

$$\sigma_k = \frac{1}{m} + \frac{1}{m(m+1)^k}.$$

If η is any positive number, it will be possible to choose a value of k for which $\sigma_k < (1/m) + \eta$. With such a value of k,

$$\nu_n \leqq \mu_{nk} \leqq A_k n^{(1/m)+\eta} \epsilon_n.$$

All this, it may now be recalled, is on the assumption that $\epsilon_n \leqq \frac{1}{4}\nu_n$. But in any case

$$\nu_n \leqq 4\epsilon_n + A_k n^{(1/m)+\eta} \epsilon_n.$$

The last relation, by the definitions of ν_n and ϵ_n, implies that

$$|f(x) - P_n(x)| = |r(x) - \pi_n(x)| \leq 5\,\varepsilon_n + A_k\, n^{(1/m)+\eta}\,\varepsilon_n$$

for $\alpha < x \leq \beta$. More concisely, since $1 \leq n^{(1/m)+\eta}$ for $n \geq 1$,

$$|f(x) - P_n(x)| \leq C n^{(1/m)+\eta}\,\varepsilon_n,$$

if $C = A_k + 5$, the factor C, like those which have preceded it in similar situations, being independent of n. The previous results with regard to convergence may be supplemented by stating:

THEOREM VIII. *If $\varrho(x)$ and $P_n(x)$ satisfy the hypotheses of Theorem* VII, *and if $\alpha_0 < \alpha < \beta < \beta_0$, the polynomials $P_n(x)$ will converge uniformly toward $f(x)$ for $\alpha \leq x \leq \beta$, if $m > 1$, and if a positive number η exists such that*

$$\lim_{\delta=0} \omega(\delta)/\delta^{(1/m)+\eta} = 0,$$

where $\omega(\delta)$ is the modulus of continuity of $f(x)$.

COROLLARY. *If $\varrho(x)$ is continuous and nowhere negative for $a \leq x \leq b$, the hypotheses with regard to $P_n(x)$ and $f(x)$ remaining unchanged, the polynomials $P_n(x)$ will converge toward $f(x)$ at all points where $\varrho(x) \neq 0$.*

For points in the interior of (a, b), this corollary supersedes the second corollary of Theorem VII, being a direct generalization of it.

There are analogous but less simple conclusions for $0 < m \leq 1$.

Generalizations of Bernstein's theorem (in appropriately modified form), leading to the extension of parts of the preceding analysis to certain cases of development in terms of characteristic functions of differential systems, have been given by Miss Carlson (Transactions of the American Mathematical Society, vol. 26 (1924), pp. 230–240, and vol. 28 (1926), pp. 435–447).

5. Polynomial approximation over an infinite interval

Another extension of the scope of the method lies in its application to problems of polynomial approximation over an

infinite interval. To be specific, let $f(x)$ be a function defined and continuous for all real values of x, and not greater in absolute value than a constant multiple of a power of x, as x becomes infinite in either direction. Among all polynomials of specified degree n, let $P_n(x)$ be the one which minimizes the integral

$$\int_{-\infty}^{\infty} e^{-x^2}\,[f(x)-P_n(x)]^2\,dx.$$

The proof of the existence of the minimum and the uniqueness of the corresponding polynomial offers no new difficulties in comparison with cases already treated, and will be omitted. (It can be based either on the construction of a sequence of orthogonal polynomials over the infinite interval with the weight-function e^{-x^2}, essentially the polynomials of Tchebychef-Hermite, or on the fundamental theorems of real analysis. From the latter point of view, the essential point in the passage to the infinite interval is merely that if any coefficient in $P_n(x)$ were large, the integral *over a finite interval* would be large, and the integral over the infinite interval would be larger still, so that the coefficients to be considered in the search for a minimum belong to a bounded domain.) It is to be shown that under suitable restrictions on $f(x)$ the polynomial $P_n(x)$ will converge everywhere to $f(x)$ as n becomes infinite, and will converge uniformly over any interval of finite extent. The aim will be to arrive at a result of this character in as straightforward a manner as possible, without regard for the greatest attainable generality.

Let γ_n be the minimum value of the integral, and let α be an arbitrary preassigned positive number. Let $p_n(x)$ be an arbitrary polynomial of the nth degree, let

$$r_n(x) = f(x)-p_n(x), \qquad \pi_n(x) = P_n(x)-p_n(x),$$

so that

$$f(x)-P_n(x) = r_n(x)-\pi_n(x),$$

and let ϵ_n be the maximum of $|r_n(x)|$ for $-(\alpha+1)\leqq x\leqq \alpha+1$. Let μ_n be the maximum of $|\pi_n(x)|$ in the same interval, attained for $x = x_0$.

By the third corollary attached to Bernstein's theorem on the derivative of a polynomial, there is a positive number H, independent of n, such that

$$|\pi_n(x) - \pi_n(x_0)| \leqq Hn\mu_n|x - x_0|^{1/2}$$

as long as x is in that one of the intervals $(-\alpha-1, 0)$, $(0, \alpha+1)$, to which x_0 belongs. This condition will be satisfied, on one side of x_0 at least, for all values of n from a certain point on, if $|x-x_0| \leqq 1/(4H^2n^2)$. So there will be an interval of length $1/(4H^2n^2)$ throughout which

$$|\pi_n(x) - \pi_n(x_0)| \leqq \frac{1}{2}\mu_n, \quad |\pi_n(x)| \geqq \frac{1}{2}\mu_n.$$

If it is supposed that $\epsilon_n \leqq \frac{1}{4}\mu_n$, then

$$|f(x) - P_n(x)| = |r_n(x) - \pi_n(x)| \geqq \frac{1}{4}\mu_n$$

throughout the same interval, whence it follows that

$$\gamma_n \geqq e^{-(\alpha+1)^2} \cdot \frac{\mu_n^2}{16} \cdot \frac{1}{4H^2n^2}, \qquad \mu_n \leqq c_1 n\gamma_n^{1/2},$$

where c_1 is independent of n. The alternative possibility that $\epsilon_n > \frac{1}{4}\mu_n$ may be left out of account until a later stage.

Now let μ'_n be the maximum of $|\pi_n(x)|$ for $-\alpha \leqq x \leqq \alpha$. It is certain that $\mu'_n < \mu_n$. Let x_1 be a point in $(-\alpha, \alpha)$ such that $|\pi_n(x_1)| = \mu'_n$. By the second corollary of Bernstein's theorem for polynomials, there is an h independent of n such that

$$|\pi'_n(x)| \leqq hn\mu_n, \quad |\pi_n(x) - \pi_n(x_1)| \leqq hn\mu_n|x - x_1|,$$

for $-\alpha \leqq x \leqq \alpha$. (It is clear that h must be positive.) The quantity $\delta_n = \mu'_n/(2hn\mu_n)$ is less than α for values of n from a certain point on, since $\mu'_n \leqq \mu_n$. One or the other of the intervals $(x_1 - \delta_n, x_1)$, $(x_1, x_1+\delta_n)$ is then contained in $(-\alpha, \alpha)$, and there is an interval of length at least δ_n throughout which

$$|\pi_n(x) - \pi_n(x_1)| \leqq \frac{1}{2}\mu'_n, \quad |\pi_n(x)| \geqq \frac{1}{2}\mu'_n.$$

If $\varepsilon_n \leq \frac{1}{4}\mu'_n$,

$$|f(x)-P_n(x)| = |r_n(x)-\pi_n(x)| \geq \frac{1}{4}\mu'_n$$

in this interval, and

$$\gamma_n \geq e^{-\alpha^2}\cdot\frac{\mu'^2_n}{16}\cdot\frac{\mu'_n}{2hn\mu_n}, \quad \mu'_n \leq c_2(n\mu_n\gamma_n)^{1/3},$$

with c_2 independent of n. The assumption that $\varepsilon_n \leq \frac{1}{4}\mu'_n$ carries with it the condition $\varepsilon_n < \frac{1}{4}\mu_n$, under which $\mu_n \leq c_1 n\gamma_n^{1/2}$, so that

$$n\mu_n\gamma_n \leq c_1 n^2\gamma_n^{3/2}, \quad \mu'_n \leq c_3 n^{2/3}\gamma_n^{1/2}.$$

If the hypothesis that $\varepsilon_n \leq \frac{1}{4}\mu'_n$ is not fulfilled, $\mu'_n < 4\varepsilon_n$ immediately. In either case

$$\mu'_n \leq c_3 n^{2/3}\gamma_n^{1/2} + 4\varepsilon_n,$$

with the reservation merely that a finite number of values of n may have had to be ruled out in the course of the proof.

As μ'_n and ε_n are upper bounds for $|\pi_n(x)|$ and $|r_n(x)|$ respectively,

$$|f(x)-P_n(x)| = |r_n(x)-\pi_n(x)| \leq c_3 n^{2/3}\gamma_n^{1/2} + 5\varepsilon_n$$

for $-\alpha \leq x \leq \alpha$. For fixed α, the polynomials $p_n(x)$ can surely be determined so that $\lim_{n\to\infty}\varepsilon_n = 0$, since $f(x)$ is continuous. *In order that $P_n(x)$ shall converge toward $f(x)$, uniformly throughout the interval $(-\alpha, \alpha)$, whatever value may be assigned to α, it is sufficient that* $\lim_{n\to\infty} n^{4/3}\gamma_n = 0$. It remains to discuss the order of magnitude of γ_n, under appropriate hypotheses on $f(x)$.

There will be occasion to use a lemma regarding the behavior of a polynomial as its argument becomes infinite. Let $p_n(x)$ be an arbitrary polynomial of the nth degree, $n \geq 1$, and let M be an upper bound for its absolute value in the interval $-1 \leq x \leq 1$. Let $X_k(x)$, as in Chapter I, be the Legendre polynomial of degree k. Then $p_n(x)$ may be written identically in the form

$$p_n(x) = A_0X_0(x) + A_1X_1(x) + \cdots + A_nX_n(x),$$

where

$$A_k = \frac{2k+1}{2}\int_{-1}^{1} p_n(x)\, X_k(x)\, dx.$$

From the last equation, together with the definition of M and the fact that $|X_k(x)| \leqq 1$ in $(-1, 1)$, it follows that

$$|A_k| \leqq (2k+1)M.$$

For $|x| \geqq 1$, on the other hand, the identity

$$X_k(x) = \frac{1}{\pi}\int_0^{\pi} [x + (x^2 - 1)^{1/2} \cos\varphi]^k\, d\varphi$$

gives

$$|X_k(x)| \leqq |2x|^k,$$

since $(x^2 - 1)^{1/2}|\cos\varphi| \leqq |x|$. Consequently, under the assumption still that $|x| \geqq 1$,

$$\begin{aligned} |p_n(x)| &\leqq M[1 + 3|2x| + 5|2x|^2 + \cdots + (2n+1)|2x|^n] \\ &\leqq M[1 + 3 + 5 + \cdots + (2n+1)]\,|2x|^n \\ &= (n+1)^2 M|2x|^n \leqq (n+n)^2 M|2x|^n = 4Mn^2|2x|^n. \end{aligned}$$

If $p_n(x)$ is a polynomial of the nth degree, as before, and if M is an upper bound for $|p_n(x)|$ in an interval $-\beta \leqq x \leqq \beta$, where β is an arbitrary positive number, then $p_n(x)$ is at the same time a polynomial of the nth degree in x/β, and for $|x| \geqq \beta$, $|x/\beta| \geqq 1$,

$$|p_n(x)| \leqq 4Mn^2|2x/\beta|^n.$$

This relation, under the hypotheses stated, constitutes the lemma in question.

Let it be assumed now that $f(x)$ has everywhere a first derivative satisfying the condition

$$|f'(x_2) - f'(x_1)| \leqq \lambda|x_2 - x_1|.$$

Let β be an arbitrary positive number, and let Theorem VII of Chapter I be applied to the approximate representation of $f(x)$ in the interval $(-\beta, \beta)$. The conclusion is that there

exists for each $n > 1$ a polynomial $p_n(x)$, of the nth degree, such that

$$|f(x) - p_n(x)| \leqq \frac{2K^2\beta^2\lambda}{n^2}$$

for $-\beta \leqq x \leqq \beta$, where K is the absolute constant that figures in Theorem I of the first chapter.

As an immediate consequence,

$$I(\beta) = \int_{-\beta}^{\beta} e^{-x^2}[f(x) - p_n(x)]^2\,dx \leqq \frac{4K^4\beta^4\lambda^2}{n^4}\int_{-\beta}^{\beta} e^{-x^2}\,dx$$
$$\leqq \frac{4K^4\beta^4\lambda^2}{n^4}\int_{-\infty}^{\infty} e^{-x^2}\,dx,$$

or, as the value of the last integral is $\pi^{1/2} < 2$,

$$I(\beta) \leqq \frac{8K^4\beta^4\lambda^2}{n^4}.$$

Furthermore, $|p_n(x)| \leqq M_1 + 2K^2\beta^2\lambda/n^2$ for $-\beta \leqq x \leqq \beta$, if M_1 is the maximum of $|f(x)|$ in the interval. Let $|f(0)| = a_0$, $|f'(0)| = a_1$; then $|f'(x)| \leqq a_1 + \lambda|x|$, $|f(x)| \leqq a_0 + a_1|x| + \frac{1}{2}\lambda x^2$, and

$$M_1 \leqq a_0 + a_1\beta + \frac{1}{2}\lambda\beta^2.$$

(Incidentally it is seen that the present hypothesis carries with it the fulfillment of the earlier requirement that $f(x)$ shall not become infinite faster than a power of x.) If $\beta \geqq 1$, as will be assumed henceforth, $M_1 \leqq (a_0 + a_1 + \frac{1}{2}\lambda)\beta^2$, and

$$|p_n(x)| \leqq c_4\beta^2$$

for $-\beta \leqq x \leqq \beta$, where c_4 is independent of β and of n, though it is different for different functions $f(x)$.

For $|x| \geqq \beta$, by the reasoning of an earlier paragraph,

$$|p_n(x)| \leqq 4c_4\beta^2 n^2|2x/\beta|^n.$$

At the same time (that is, for $|x| \geqq \beta \geqq 1$)

$$|f(x)| \leqq a_0 + a_1|x| + \frac{1}{2}\lambda x^2 \leqq \left(a_0 + a_1 + \frac{1}{2}\lambda\right)x^2,$$

and by further enlargement of the right-hand member, with $a_2 = a_0 + a_1 + \frac{1}{2}\lambda$ for abbreviation,

$$|f(x)| \leqq a_2 x^2 = \frac{1}{4} a_2 \beta^2 (2x/\beta)^2 \leqq \frac{1}{4} a_2 \beta^2 n^2 (2x/\beta)^2$$
$$\leqq \frac{1}{4} a_2 \beta^2 n^2 |2x/\beta|^n;$$

it has already been assumed that $n \geqq 2$. So, with a new multiplier c_5, independent of β and of n,

$$|f(x) - p_n(x)| \leqq \left(4 c_4 + \frac{1}{4} a_2\right) \beta^2 n^2 |2x/\beta|^n,$$
$$[f(x) - p_n(x)]^2 \leqq c_5 \beta^4 n^4 |2x/\beta|^{2n}$$

for $|x| \geqq \beta$.

This means that

$$\int_\beta^\infty e^{-x^2} [f(x) - p_n(x)]^2 \, dx \leqq c_5 \beta^4 n^4 \left(\frac{2}{\beta}\right)^{2n} \int_\beta^\infty x^{2n} e^{-x^2} \, dx;$$

the quantity on the right is an upper bound also for the integral from $-\infty$ to $-\beta$. Since $\beta \geqq 1$,

$$\int_\beta^\infty x^{2n} e^{-x^2} \, dx \leqq \int_\beta^\infty x^{2n} e^{-x^2} x \, dx = \frac{1}{2} \int_{\beta^2}^\infty y^n e^{-y} \, dy$$
$$\leqq \frac{1}{2} \int_0^\infty y^n e^{-y} \, dy = \frac{1}{2} \Gamma(n+1) = \frac{1}{2} n! = \frac{1}{2} n (n-1)!$$

Without reference to more elaborate approximations to the value of the factorial, it is seen at once that

$$\log k = \int_k^{k+1} \log k \, dt < \int_k^{k+1} \log t \, dt, \quad k = 1, 2, \cdots,$$
$$\log [(n-1)!] = \log 1 + \log 2 + \cdots + \log (n-1)$$
$$< \int_1^n \log t \, dt = n \log n - n + 1,$$
$$(n-1)! < e (n/e)^n.$$

Consequently

$$\int_\beta^\infty e^{-x^2} [f(x) - p_n(x)]^2 \, dx \leqq \frac{1}{2} c_5 e \beta^4 n^5 \left(\frac{2}{\beta}\right)^{2n} \left(\frac{n}{e}\right)^n$$
$$= \frac{1}{2} c_5 e \beta^4 n^5 \left(\frac{4n}{\beta^2 e}\right)^n,$$

and

$$g_n = \int_{-\infty}^{\infty} e^{-x^2} [f(x) - p_n(x)]^2 \, dx$$

$$\leqq \frac{8\, K^4 \beta^4 \lambda^2}{n^4} + c_5\, e \beta^4 n^5 \left(\frac{4\, n}{\beta^2 e}\right)^n .$$

The number β has been arbitrary hitherto, except for the requirement that it be positive and not less than 1; the last relation is valid with the insertion of any such value of β. *Now let* $\beta = 2\, n^{1/2}$. This makes

$$g_n \leqq \frac{128\, K^4 \lambda^2}{n^2} + 16\, c_5\, e n^7 e^{-n} .$$

As $n^9 e^{-n}$ approaches zero when n becomes infinite, $n^7 e^{-n}$ does not exceed a constant multiple of $1/n^2$ for $n \geqq 2$, while, on the other hand, g_n is not less than the minimum value γ_n. Therefore

$$\gamma_n \leqq g_n \leqq \frac{c_6}{n^2} ,$$

with c_6 independent of n.

The condition $\lim_{n=\infty} n^{4/3} \gamma_n = 0$ being satisfied, $P_n(x)$ *converges uniformly toward* $f(x)$ *throughout any finite interval,* under the hypotheses stated.

As in the case of the Fourier and Legendre series, the method under discussion, when applied merely to the classical problem, yields little if anything that is new, and misses much that is well known. Clearly, however, the treatment admits a variety of generalizations, which remain open for further investigation. To mention just one, which calls for no additional labor, the reasoning applies without material change if the weight function e^{-x^2} is replaced by any positive continuous function which is never greater than a constant multiple of e^{-x^2}. The theory thus suggested has been developed at some length in a thesis, as yet unpublished, by J. M. Earl. Theorems on degree of approximation over an infinite interval, without reference to the particular method of this chapter, have been published by W. E. Milne in vol. 31 of the Transactions of the American Mathematical Society.

CHAPTER IV

TRIGONOMETRIC INTERPOLATION

1. Fundamental formulas of trigonometric interpolation

This chapter is concerned with certain striking analogies, both formal and analytical, between the theory of interpolation by means of trigonometric sums and that of Fourier series. The case of polynomial interpolation will be left out of consideration for the most part, since the analogies there, when the points used for interpolating are equally spaced, are rather with Taylor's series than with those of Fourier and Legendre. Certain extensions to the case of interpolation by means of Sturm-Liouville sums have been carried through by C. M. Jensen (Transactions of the American Mathematical Society, vol. 29 (1927), pp. 54–79).

Let $t_0, t_1, \cdots, t_{2n}$ be $2n+1$ distinct numbers contained in an interval of length 2π, for definiteness that from 0 to 2π, inclusive of the left-hand and exclusive of the right-hand end point. Let $y_0, y_1, \cdots, y_{2n}$ be any $2n+1$ real numbers, distinct or not. Let the problem be proposed of determining a trigonometric sum $T_n(x)$, of the nth order, to satisfy the conditions

$$T_n(t_r) = y_r, \qquad r = 0, 1, \cdots, 2n.$$

The sum $T_n(x)$ has $2n+1$ coefficients, on which $2n+1$ conditions are imposed. With the notation

$$\begin{aligned} T_n(x) = a_0 + a_1 \cos x + \cdots + a_n \cos nx \\ + b_1 \sin x + \cdots + b_n \sin nx, \end{aligned}$$

the conditions to be satisfied are given explicitly by the $2n+1$ equations

$$
\begin{aligned}
a_0 + a_1 \cos t_0 + b_1 \sin t_0 + \cdots + a_n \cos n t_0 + b_n \sin n t_0 &= y_0 , \\
a_0 + a_1 \cos t_1 + b_1 \sin t_1 + \cdots + a_n \cos n t_1 + b_n \sin n t_1 &= y_1 , \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
a_0 + a_1 \cos t_{2n} + b_1 \sin t_{2n} + \cdots + a_n \cos n t_{2n} + b_n \sin n t_{2n} &= y_{2n} ,
\end{aligned}
$$

which are linear in the a's and b's. One is confronted by the problem of showing, directly or indirectly, that the determinant of these equations, the determinant

$$
\begin{vmatrix}
1 & \cos t_0 & \sin t_0 & \cdots & \cos n t_0 & \sin n t_0 \\
1 & \cos t_1 & \sin t_1 & \cdots & \cos n t_1 & \sin n t_1 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
1 & \cos t_{2n} & \sin t_{2n} & \cdots & \cos n t_{2n} & \sin n t_{2n}
\end{vmatrix},
$$

is different from zero.

If α and β are any two numbers, the product

$$\sin \frac{1}{2}(x-\alpha) \sin \frac{1}{2}(x-\beta)$$

is a trigonometric sum of the first order in x, being identically equal to

$$
\frac{1}{2}\left[\cos\left(\frac{1}{2}\alpha - \frac{1}{2}\beta\right) - \cos\left(x - \frac{1}{2}\alpha - \frac{1}{2}\beta\right)\right]
$$
$$= A_0 + A_1 \cos x + B_1 \sin x ,$$

with $A_0 = \frac{1}{2}\cos(\frac{1}{2}\alpha - \frac{1}{2}\beta)$, $A_1 = -\frac{1}{2}\cos(\frac{1}{2}\alpha + \frac{1}{2}\beta)$, $B_1 = -\frac{1}{2}\sin(\frac{1}{2}\alpha + \frac{1}{2}\beta)$. In the expression

$$
\varphi_k(x) = \frac{\sin \frac{1}{2}(x-t_0)\cdots\sin\frac{1}{2}(x-t_{k-1})\sin\frac{1}{2}(x-t_{k+1})\cdots\sin\frac{1}{2}(x-t_{2n})}{\sin \frac{1}{2}(t_k-t_0)\cdots\sin\frac{1}{2}(t_k-t_{k-1})\sin\frac{1}{2}(t_k-t_{k+1})\cdots\sin\frac{1}{2}(t_k-t_{2n})},
$$

the $2n$ factors of the numerator can be combined in pairs to give n expressions, each of which is a trigonometric sum of the first order, and the product of these is a trigonometric sum of the nth order; the denominator is a constant which is not zero, since the difference of any two of the numbers t_r is by hypothesis different from zero and numerically less than 2π. It is apparent on inspection that $\varphi_k(t_r) = 0$ for $r \neq k$, while $\varphi_k(t_k) = 1$. *Consequently the function* $T_n(x)$ *defined by the formula*

$$T_n(x) = \sum_{k=0}^{2n} y_k \varphi_k(x)$$

is a trigonometric sum of the nth order such that $T_n(t_r) = y_r$, $r = 0, 1, \cdots, 2n$. The linear equations above accordingly have at least one solution, for any assigned values of the y's. But if the determinant of the equations were zero, there would be values of the right-hand members for which there would be no solution. *It is certain therefore that the determinant is different from zero.* Whatever values are given to the y's, the proposed problem has one and just one solution. In particular, if all the y's are zero, the obvious solution in which all the a's and b's are zero is the only one. *A trigonometric sum of the nth order which vanishes at* $2n+1$ *distinct points in a period is identically zero; two trigonometric sums of the nth order which coincide in value at* $2n+1$ *distinct points of a period are identically equal.*

It is easy now to supply a proof of a fact which was previously assumed as known, in connection with the proof of Bernstein's theorem, namely that a trigonometric sum of the nth order vanishes identically if it has $2n$ distinct roots in a period, one of which is a double root. Let $T_n(x)$ be a sum of the nth order which takes on the value 0 for $2n$ distinct points of a period, say for $x = t_1, t_2, \cdots, t_{2n}$, but which does not vanish identically. Let t_0 be a point of the same period interval, distinct from $t_1, \cdots, t_{2n}$, and let $y_0 = T_n(t_0)$: it is certain that $y_0 \neq 0$, since $T_n(x)$ can not have $2n+1$ distinct roots in the interval. The expression

$$y_0 \frac{\sin\frac{1}{2}(x-t_1)\sin\frac{1}{2}(x-t_2)\cdots\sin\frac{1}{2}(x-t_{2n})}{\sin\frac{1}{2}(t_0-t_1)\sin\frac{1}{2}(t_0-t_2)\cdots\sin\frac{1}{2}(t_0-t_{2n})}$$

s a trigonometric sum of the nth order which takes on the value y_0 for $x = t_0$, and vanishes for $x = t_1, \cdots, t_{2n}$, and so must be identical with $T_n(x)$. The derivative of this expression is easily calculated explicitly; it is

$$T_n'(x) = \frac{1}{2} y_0 \sum_{k=1}^{2n} \frac{\sin\frac{1}{2}(x-t_1)\cdots\cos\frac{1}{2}(x-t_k)\cdots\sin\frac{1}{2}(x-t_{2n})}{\sin\frac{1}{2}(t_0-t_1)\cdots\sin\frac{1}{2}(t_0-t_k)\cdots\sin\frac{1}{2}(t_0-t_{2n})}.$$

For $x = t_r$, if r has any of the values $1, \cdots, 2n$, all the terms of the summation vanish, except the one for which $k = r$, and that becomes

$$D_r = \frac{\sin\frac{1}{2}(t_r - t_1) \cdots \qquad 1 \qquad \cdots \sin\frac{1}{2}(t_r - t_{2n})}{\sin\frac{1}{2}(t_0 - t_1) \cdots \sin\frac{1}{2}(t_0 - t_r) \cdots \sin\frac{1}{2}(t_0 - t_{2n})},$$

which is certainly different from zero, so that $T_n'(t_r) = \frac{1}{2} y_0 D_r \neq 0$; the sum $T_n(x)$, assumed not to vanish identically, can not have a double root at any of the points $t_1, \cdots, t_{2n}$.

To return to the problem of interpolation, *let the points t_r from now on be supposed equally spaced over a period.* As a matter of general notation, if m is any positive integer and r any real integer, positive, negative, or zero, let

$$t_r = 2r\pi/m.$$

The discussion hitherto (apart from the digression of the last paragraph) is applicable on the assumption that m is odd, $m = 2n+1$; there will be occasion subsequently to consider even values of m as well. *Throughout the remainder of this chapter, in the absence of express indication to the contrary, the sign $\sum$ will be understood to refer to summation with respect to the index r, from $r = 0$ to $r = m-1$*, or what comes to the same thing, as all the functions considered will be of period 2π, over any m successive values of r.

It is a fundamental fact, whether m is odd or even, that

$$\sum \sin k t_r = 0$$

if k is any integer, and that

$$\sum \cos k t_r = 0$$

if k is an integer not divisible by m, while if k is a multiple of m (and in particular if $k = 0$) it is evident that

$$\sum \cos k t_r = m,$$

since in this case each term of the summation is equal to 1.

The truth of the statement with regard to $\sum \sin k t_r$ is

apparent from considerations of symmetry. The term $\sin k t_0$ is equal to 0. Since

$$k t_{m-r} = 2k(m-r)\pi/m = 2k\pi - kt_r,$$

each term corresponding to an index r between 0 and $\frac{1}{2}m$ is paired with a numerically equal term of opposite sign. If m is odd, all the terms are thus accounted for; if m is even, the term which remains by itself is $\sin k t_{m/2} = \sin k\pi = 0$.

In connection with the cosine sum it is to be recalled that

$$\frac{1}{2} + \cos u + \cos 2u + \cdots + \cos nu = \frac{\sin(n+\frac{1}{2})u}{2\sin\frac{1}{2}u}.$$

(The right-hand member is understood to be defined by continuity at points where the denominator vanishes.) If $u = kt_1$, then $ru = kt_r$, and

$$\frac{1}{2} + \cos kt_1 + \cos kt_2 + \cdots + \cos kt_n = \frac{\sin(n+\frac{1}{2})kt_1}{2\sin\frac{1}{2}kt_1}.$$

But by the symmetry pointed out in the preceding paragraph, the terms on the left are respectively the same as those of the sum

$$\frac{1}{2} + \cos kt_{m-1} + \cos kt_{m-2} + \cdots + \cos kt_{m-n}.$$

If $m = 2n+1$, and if k is not divisible by m, addition of the two sums gives

$$\sum \cos kt_r = \frac{\sin(n+\frac{1}{2})kt_1}{\sin\frac{1}{2}kt_1},$$

and in the last expression the numerator is zero, since $t_1 = 2\pi/m$ and $n+\frac{1}{2} = \frac{1}{2}m$, so that $(n+\frac{1}{2})kt_1 = k\pi$, while the denominator is different from zero. If $m = 2n$, the term $\cos kt_n = \cos nkt_1$ occurs in each sum; combination of the two, with subtraction of the redundant term, gives

$$\begin{aligned}\sum \cos kt_r &= \frac{\sin(n+\frac{1}{2})kt_1}{\sin\frac{1}{2}kt_1} - \cos nkt_1\\ &= \frac{\sin nkt_1 \cos\frac{1}{2}kt_1 + \cos nkt_1 \sin\frac{1}{2}kt_1}{\sin\frac{1}{2}kt_1} - \cos nkt_1\\ &= \sin nkt_1 \cot\frac{1}{2}kt_1 = \sin k\pi \cot(k\pi/m) = 0,\end{aligned}$$

the assumption being still that k is not divisible by m.

The proposition of which the preceding lines give an analytical proof is almost evident geometrically, though a precise formulation of the geometrical argument requires a little attention to detail. The pairs of numbers $(\cos t_r, \sin t_r)$, $r = 0, 1, \cdots, m-1$, are the coördinates of the vertices of a regular polygon of m sides having its center at the origin, and the quantities $(1/m)\sum \cos t_r$, $(1/m)\sum \sin t_r$, being the coördinates of the center of gravity of these points, must evidently be zero. The more general fact of the vanishing of $\sum \cos kt_r$ and $\sum \sin kt_r$ is then obtained by consideration of the various possibilities as to the existence of common factors of k and m.

In the following statements, let p and q be integers subject to the restrictions $0 \leqq p \leq \frac{1}{2}m$, $0 < q \leq \frac{1}{2}m$. Then, in the identities

$$\cos px \cos qx = \frac{1}{2}[\cos(p-q)x + \cos(p+q)x],$$

$$\sin px \sin qx = \frac{1}{2}[\cos(p-q)x - \cos(p+q)x],$$

$$\cos px \sin qx = \frac{1}{2}[\sin(p+q)x - \sin(p-q)x],$$

$p-q$ and $p+q$ are integers numerically less than m, except that $p+q = m$ if m is even and $p = q = \frac{1}{2}m$, so that

$$\begin{array}{ll}
\sum \cos pt_r \cos qt_r = 0 & \text{if } p \neq q, \\
\sum \sin pt_r \sin qt_r = 0 & \text{if } p \neq q, \\
\sum \cos pt_r \sin qt_r = 0 & \text{for all } p \text{ and } q, \\
\sum \cos^2 pt_r = \sum \sin^2 pt_r = \frac{1}{2}m & \text{if } 0 < p < \frac{1}{2}m, \\
\sum \cos^2 pt_r = m, \quad \sum \sin^2 pt_r = 0 & \text{if } p = 0 \text{ or } \frac{1}{2}m.
\end{array}$$

Until the contrary is stated, let it be supposed now that m is odd, $m = 2n+1$. The equations for the interpolating coefficients given in the second paragraph of the chapter can be solved explicitly, under the present hypothesis of equal spacing of the points t_r. It is known in advance that

they have one and just one solution. If the $2n+1$ equations are added as they stand, the left-hand member in the result reduces to $(2n+1)a_0$, so that

$$a_0 = \frac{1}{2n+1}\sum y_r.$$

(As to the convention with regard to the meaning of the sign $\sum$, it is to be understood here that the summation is performed specifically over the range $r = 0, 1, \cdots, 2n$, or else that y_r is defined outside this range as a periodic function of the index by the prescription that $y_{r+2n+1} = y_r$.) To determine a_k, $k > 0$, let the $(r+1)$th equation be multiplied by $\cos kt_r$, $r = 0, 1, \cdots, 2n$. On addition of the equations thus obtained, the left-hand member in the sum becomes $a_k \sum \cos^2 kt_r = \frac{1}{2} m a_k = \frac{1}{2}(2n+1)a_k$, whence

$$a_k = \frac{2}{2n+1}\sum y_r \cos kt_r.$$

Similarly,

$$b_k = \frac{2}{2n+1}\sum y_r \sin kt_r.$$

If the notation is modified to the extent of representing the constant term in the interpolating sum by $a_0/2$, instead of a_0, the general formula for a_k gives the correct value of a_0 also. It will be understood henceforth that the notation is adjusted in this way, and the interpolating sum for equally spaced points t_r will be denoted by $S_n(x)$, so that

$$\begin{aligned} S_n(x) = \frac{1}{2}a_0 + a_1 \cos x + \cdots + a_n \cos nx \\ + b_1 \sin x + \cdots + b_n \sin nx. \end{aligned}$$

the a's and b's being given by the two preceding equations for $k = 0, 1, \cdots, n$. The resemblance to the Fourier coefficients is apparent.

Even if it had not been known *a priori* that the equations have a unique solution, that fact would be an immediate consequence of the work of the last paragraph. For the

work shows that if the equations have a solution, it must be the one indicated; that is, they can not have more than one solution for any given set of y's. But if the determinant of the equations were zero, there would be values of the y's for which there would be infinitely many solutions. This new proof of the non-vanishing of the determinant, however, depends essentially on the assumption that the t's are equidistant, and so is less general than the one previously given.

The formal resemblance to the case of Fourier series is further borne out by substitution of the values of the a's and b's in the expression defining $S_n(x)$, and rearrangement of the result by means of the identity for a sum of cosines:

$$\begin{aligned} S_n(x) &= \frac{2}{2n+1}\sum y_r\left[\frac{1}{2}+\sum_{k=1}^{n}(\cos kt_r\cos kx+\sin kt_r\sin kx)\right] \\ &= \frac{2}{2n+1}\sum y_r\left[\frac{1}{2}+\sum_{k=1}^{n}\cos k(t_r-x)\right] \\ &= \frac{1}{2n+1}\sum y_r\frac{\sin(n+\frac{1}{2})(t_r-x)}{\sin\frac{1}{2}(t_r-x)}. \end{aligned}$$

Incidentally, the correctness of the last expression as a solution of the problem of interpolation can be verified directly, by substituting $x=t_q$, for any particular value of q. For $t_r-t_q=2(r-q)\pi/(2n+1)$, $\sin(n+\frac{1}{2})(t_r-t_q)=\sin(r-q)\pi=0$, and all the terms of the summation vanish except the single one with a vanishing denominator, while the limiting value of the quotient of sines in that term is $2n+1$, so that $S_n(t_q)=y_q$.

Now let the problem be changed by supposing that m is even, $m=2n$. Corresponding to the $2n$ abscissas $t_0,\cdots,t_{2n-1}$, let $2n$ numbers $y_0,\cdots,y_{2n-1}$ be given, and let y_r be defined for other values of r so that $y_{r+2n}=y_r$. Consider the question of the existence of a trigonometric sum $T_n(x)$, of the nth order, represented for the moment by the notation of the second paragraph of the chapter, to satisfy the conditions $T_n(t_r)=y_r$, $r=0,1,\cdots,2n-1$. It is not to be expected that the problem will have a definite solution, since

only $2n$ conditions are imposed on the $2n+1$ coefficients. The precise nature of the indeterminacy is brought out by going through the formal manipulation of the equations. Written out at length, these are as follows:

$$a_0 + a_1 \cos t_0 + b_1 \sin t_0 + \cdots + a_n \cos n t_0 + b_n \sin n t_0 = y_0,$$

$$a_0 + a_1 \cos t_1 + b_1 \sin t_1 + \cdots + a_n \cos n t_1 + b_n \sin n t_1 = y_1,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_0 + a_1 \cos t_{2n-1} + b_1 \sin t_{2n-1} + \cdots + a_n \cos n t_{2n-1} + b_n \sin n t_{2n-1} = y_{2n-1}.$$

Direct addition of them gives

$$2na_0 = \sum y_r, \qquad a_0 = (\sum y_r)/2n.$$

Multiplied respectively by $\cos kt_0$, $\cos kt_1, \cdots, \cos kt_{2n-1}$, for any value of k from 1 to $n-1$ inclusive, and then added, they yield

$$a_k = (1/n) \sum y_r \cos k t_r,$$

and similarly

$$b_k = (1/n) \sum y_r \sin k t_r.$$

So far there is no ambiguity, and no formal difference from the case previously treated, except the replacement of $2n+1$ by $2n$. When the $(r+1)$th equation is multiplied by $\cos nt_r$, there is a difference because in the present case $n = \frac{1}{2}m$, and $\sum \cos^2 nt_r$ is equal to $m = 2n$, instead of $\frac{1}{2}m$. The determination is still perfectly definite, however:

$$a_n = (\sum y_r \cos n t_r)/(2n).$$

But as $t_r = r\pi/n$, the numbers $\sin n t_r$ are all zero, and the use of these quantities as multipliers does not lead to any determination of b_n at all.

It is now apparent how the problem of interpolation with an even number of equidistant abscissas can be formulated so as to have a determinate solution. If $S_n(x)$ denotes an expression of the form

$$\frac{1}{2}a_0 + a_1 \cos x + \cdots + a_{n-1} \cos (n-1)x + \frac{1}{2} a_n \cos nx$$
$$+ b_1 \sin x + \cdots + b_{n-1} \sin (n-1)x,$$

the conditions $S_n(t_r) = y_r$ are expressed by $2n$ linear equations for the $2n$ coefficients. The solution of these equations, if they have a solution, is given uniquely by the formulas

$$a_k = (1/n) \sum y_r \cos kt_r, \qquad b_k = (1/n) \sum y_r \sin kt_r,$$

for $k = 0, 1, \cdots, n$. But if the determinant of the system were zero, there would be values of the y's for which more than one solution would exist. So the value of the determinant is certainly different from zero, and the problem has the unique solution indicated. As $\sin nt_r = 0$ for all values of r, the expression

$$S_n(x) + b_n \sin nx,$$

with any value of b_n, likewise takes on the value y_r when $x = t_r$; the indeterminacy of the problem as first proposed consists merely in the complete indeterminacy of the coefficient b_n. The subsequent discussion will relate to the sum $S_n(x)$ as defined at the beginning of this paragraph, without the term in $\sin nx$.

For deriving a concise expression for $S_n(x)$, to be sure, it is convenient as a matter of form to include a term $\frac{1}{2}b_n \sin nx$, the coefficient b_n being defined by the general formula for b_k, with $k = n$; as each term of the summation vanishes, this gives $b_n = 0$ automatically. Then it appears that

$$S_n(x) = \frac{1}{n} \sum y_r \left[\frac{1}{2} + \sum_{k=1}^{n-1} \cos k(t_r - x) + \frac{1}{2} \cos n(t_r - x) \right],$$

or, as

$$\frac{1}{2} + \cos u + \cdots + \cos (n-1)u + \frac{1}{2} \cos nu$$
$$= \frac{1}{2} \sin nu \cot \frac{1}{2} u,$$

by subtraction of $\frac{1}{2} \cos nu$ from the identity previously used,

$$S_n(x) = \frac{1}{2n} \sum y_r \sin n(t_r - x) \cot \frac{1}{2} (t_r - x).$$

As in the case of an odd number of points, substitution of $x = t_q$ gives a direct verification of the validity of the formula for purposes of interpolation. The analogy with the partial sum of a Fourier series, however, is not superficially so much in evidence as before.

2. Convergence and degree of convergence under hypotheses of continuity over entire period

For the discussion of the analytical properties of the interpolating sums $S_n(x)$, particularly questions of convergence as n becomes infinite, let the given numbers y_r be values of a function having specified properties: let $f(x)$ be a given function of period 2π, and let $y_r = f(t_r)$. Then $S_n(x)$ coincides in value with $f(x)$ at the points t_r, and the question is to what extent $S_n(x)$ furnishes an approximation to $f(x)$ at intermediate points.

In the first place, let $f(x)$ be a bounded function having M as an upper bound for its absolute value. Then a corresponding upper bound can be assigned for $|S_n(x)|$. Let $m = 2n+1$. If x has one of the values t_r, $|S_n(x)| = |f(t_r)| < M$. If x is not one of the numbers t_r, let t_R be that one of these numbers which is nearest to x, or one of the two nearest, if two are equally near. Then

$$t_R - \frac{\pi}{2n+1} \leqq x \leqq t_R + \frac{\pi}{2n+1}.$$

Because of the periodicity of the functions concerned, as has been noted already, the summation with regard to r in the formula representing $S_n(x)$ can be extended over any $2n+1$ successive values of r, and in particular from $R-n$ to $R+n$ inclusive. Accordingly

$$|S_n(x)| = \frac{1}{2n+1}\left|\sum_{r=R-n}^{R+n} f(t_r)\frac{\sin(n+\frac{1}{2})(t_r-x)}{\sin\frac{1}{2}(t_r-x)}\right|$$
$$\leqq \frac{M}{2n+1}\sum_{r=R-n}^{R+n}\left|\frac{\sin(n+\frac{1}{2})(t_r-x)}{\sin\frac{1}{2}(t_r-x)}\right|.$$

The absolute value of the last expression in bars never exceeds $2n+1$, as may be seen from its interpretation as a sum of cosines. This observation will be sufficient as far as the three terms corresponding to the indices $r = R-1, R, R+1$ are concerned. The numbers $t_{R+2}-x$, $t_{R+3}-x$, $\cdots$, $t_{R+n}-x$ are respectively greater than $2\pi/(2n+1)$, $4\pi/(2n+1)$, $\cdots$ $\cdots$, $2(n-1)\pi/(2n+1)$; the same thing may be said of the sequence of numbers $|t_{R-2}-x|$, $|t_{R-3}-x|$, $\cdots$, $|t_{R-n}-x|$; and for all the values of r in question, $|t_r - x| < \pi$, so that $|\sin\frac{1}{2}(t_r - x)| > (1/\pi)|t_r - x|$. Hence

$$\sum_{r=R-n}^{R+n} \left| \frac{\sin(n+\frac{1}{2})(t_r - x)}{\sin\frac{1}{2}(t_r - x)} \right|$$
$$\leq 3(2n+1) + \left(\sum_{r=R-n}^{R-2} + \sum_{r=R+2}^{R+n}\right) \left| \frac{\sin(n+\frac{1}{2})(t_r - x)}{\sin\frac{1}{2}(t_r - x)} \right|$$
$$\leq 3(2n+1) + \left(\sum_{r=R-n}^{R-2} + \sum_{r=R+2}^{R+n}\right) \frac{1}{|\sin\frac{1}{2}(t_r - x)|}$$
$$\leq 3(2n+1) + \pi\left(\sum_{r=R-n}^{R-2} + \sum_{r=R+2}^{R+n}\right) \frac{1}{|t_r - x|}$$
$$< 3(2n+1) + (2n+1)\sum_{j=1}^{n-1}\frac{1}{j} = 4(2n+1) + (2n+1)\sum_{j=2}^{n-1}\frac{1}{j}$$
$$\leq 4(2n+1) + (2n+1)\int_1^{n-1}\frac{du}{u} < 4(2n+1) + (2n+1)\log(n-1)$$
$$< (2n+1)(4+\log n),$$

and

$$|S_n(x)| < M(4+\log n).$$

The last parenthesis does not exceed a constant multiple of $\log n$, for $n > 2$, and $|S_n(x)|$ does not exceed a constant multiple of $M \log n$.

Similar reasoning, with minor differences of detail, applies when $m = 2n$, the sum $S_n(x)$ then being given by the formula of the third paragraph preceding. The conclusion may be stated comprehensively for both cases as

LEMMA I. *If $f(x)$ is a function of period 2π satisfying the condition that*

$$|f(x)| \leq M$$

for all values of x, and if $S_n(x)$ is the interpolating sum of the nth order for $f(x)$ corresponding to the subdivision of a period either into $2n+1$ or into $2n$ equal parts, then

$$|S_n(x)| \leqq CM \log n$$

for $n > 1$, where C is an absolute constant.

The representation of the constant by the notation used in the lemma preceding Theorem IX of Chapter I does not imply that the constants are the same, though of course both lemmas could be stated together, with a single symbol to represent the larger of the two constants.

For the application of the lemma, it is to be noted that $|f(x) - S_n(x)| \leqq M + CM \log n$, which is likewise not greater than a constant multiple of $M \log n$, say $BM \log n$; that the interpolating expression corresponding to the sum of two given functions is the sum of the interpolating expressions constructed for the two functions separately, and the error of the sum is the sum of the errors; and that if $m = 2n+1$ the interpolating sum $S_n(x)$ formed for a function $f(x)$ which is itself a trigonometric sum $T_n(x)$ of the nth order is identical with $T_n(x)$, since by reason of the interpolating property $S_n(x)$ and $T_n(x)$ are trigonometric sums of the nth order coinciding in value at $2n+1$ distinct points of a period. Taken together with the lemma, these observations yield

THEOREM I. *If $f(x)$ is a function of period 2π, if $S_n(x)$ is the interpolating sum of the nth order for $f(x)$ corresponding to the subdivision of a period into $2n+1$ equal parts, $n > 1$, and if there exists a trigonometric sum $T_n(x)$, of the nth order, such that*

$$|f(x) - T_n(x)| \leqq \varepsilon_n$$

for all values of x, then, for all values of x,

$$|f(x) - S_n(x)| \leqq B\varepsilon_n \log n,$$

where B is an absolute constant.

When $m = 2n$, the interpolating sum of the nth order, as defined above, for a trigonometric sum $T_n(x)$ of the nth order

is not in general the same as $T_n(x)$, since the interpolating sum does not contain any term in $\sin nx$. But the interpolating sum $S_{n+1}(x)$ of order $n+1$, obtained by taking $m = 2n+2$, is identical with $T_n(x)$. For the coefficient a_{n+1} in $S_{n+1}(x)$, given by the expression

$$[1/(n+1)] \sum T_n(t_r) \cos (n+1) t_r,$$

reduces to zero, since each of the sines and cosines in $T_n(x)$, through $\sin nx$ and $\cos nx$, is orthogonal to $\cos (n+1)x$ for summation over the finite range in question; and it follows that $S_{n+1}(x)$ and $T_n(x)$ are trigonometric sums *of the nth order* agreeing in value at $2n+2$ distinct points of a period. The above statement may therefore be supplemented as follows:

THEOREM I (*continued*). *If $S_{n+1}(x)$ is the interpolating sum of order $n+1$ corresponding to the subdivision of a period into $2n+2$ equal parts, the hypotheses remaining otherwise unchanged, then*

$$|f(x) - S_{n+1}(x)| \leqq B\epsilon_n \log (n+1),$$

where B has the same value as before.

As $\log (n+1) < 2 \log n$ for $n > 1$, the right-hand member may be replaced by $2B\epsilon_n \log n$. As an alternative, n may be replaced by $n-1$, to give

$$|f(x) - S_n(x)| \leqq B\epsilon_{n-1} \log n$$

for $m = 2n$.

Like Theorem IX of Chapter I, the present Theorem I can be combined with Theorems I-IV of Chapter I to give a succession of more specific results. The cases of an odd number and of an even number of subdivisions can be covered by a single formulation each time, by virtue of the observation that $1/(n-1) \leqq 2/n$ and $\omega[2\pi/(n-1)] \leqq \omega(4\pi/n) \leqq 2\omega(2\pi/n)$, for $n > 1$. The symbol A is merely a notation for the largest of what would be obtained in the first instance as a finite number of different constants, and A_p similarly:

COROLLARY I. *If*

$$|f(x_2) - f(x_1)| \leqq \lambda |x_2 - x_1|,$$

for all values of x_1 *and* x_2, λ *being a constant, then*

$$|f(x) - S_n(x)| \leq \frac{A\lambda \log n}{n}$$

for $m = 2n+1$ *and for* $m = 2n$.

COROLLARY II. *If* $f(x)$ *is continuous with modulus of continuity* $\omega(\delta)$,

$$|f(x) - S_n(x)| \leq A\omega\left(\frac{2\pi}{n}\right)\log n$$

for $m = 2n+1$ *and for* $m = 2n$.

COROLLARY IIa. *The interpolating sum* $S_n(x)$ *converges uniformly to the value* $f(x)$, *as* m *becomes infinite through odd or even values (or both), if* $f(x)$ *has a modulus of continuity* $\omega(\delta)$ *such that* $\lim_{\delta=0}\omega(\delta)\log\delta = 0$.

COROLLARY III. *If* $f(x)$ *has a pth derivative* $f^{(p)}(x)$ *such that*

$$|f^{(p)}(x_2) - f^{(p)}(x_1)| \leq \lambda |x_2 - x_1|$$

for all values of x_1 *and* x_2, λ *being a constant, then*

$$|f(x) - S_n(x)| \leq \frac{A_p\lambda \log n}{n^{p+1}},$$

for $m = 2n+1$ *and for* $m = 2n$.

COROLLARY IV. *If* $f(x)$ *has everywhere a continuous pth derivative with modulus of continuity* $\omega(\delta)$,

$$|f(x) - S_n(x)| \leq \frac{A_p}{n^p}\,\omega\left(\frac{2\pi}{n}\right)\log n.$$

for $m = 2n+1$ *and for* $m = 2n$.

In each of these statements, the conclusion holds for all values of x, *and for all values of* $n \geq 2$; *the coefficient* A *is an absolute constant, while* A_p *depends only on* p.

3. Convergence under hypothesis of continuity over part of a period

The next proposition is analogous to a result obtained in the second and third paragraphs of Chapter II:

LEMMA II. *If $f(x)$ is a function of period 2π, bounded and integrable in the sense of Riemann,*

$$\lim (1/m) \sum f(t_r) \cos n t_r = \lim (1/m) \sum f(t_r) \sin n t_r = 0,$$

as m becomes infinite through odd values ($m = 2n+1$) or even values ($m = 2n$) or both.

There is a noteworthy difference in the hypothesis imposed on $f(x)$; it is in the nature of the case no longer sufficient to assume merely that the function is summable, or summable with its square, since an enumerable set of very large values of $f(x)$, while not affecting the definite integrals of the earlier chapter, might throw the present sums entirely out of proportion.

Suppose first that $m = 2n+1$. Let the summation be extended for definiteness from $r = 0$ to $r = 2n$. Since $t_r = 2r\pi/(2n+1)$,

$$n t_r = \frac{2rn\pi}{2n+1} = \frac{2r(n+\frac{1}{2})\pi - r\pi}{2n+1} = r\pi - \frac{r\pi}{2n+1},$$

$$\cos n t_r = (-1)^r \cos \frac{r\pi}{2n+1},$$

$$\frac{1}{2n+1} \sum_{r=0}^{2n} f(t_r) \cos n t_r$$

$$= \frac{1}{2\pi} \sum_{j=0}^{n} \frac{2\pi}{2n+1} f\left(\frac{4j\pi}{2n+1}\right) \cos \frac{2j\pi}{2n+1}$$

$$- \frac{1}{2\pi} \sum_{j=1}^{n} \frac{2\pi}{2n+1} f\left(\frac{2(2j-1)\pi}{2n+1}\right) \cos \frac{(2j-1)\pi}{2n+1}.$$

When n becomes infinite, the first sum in the last member approaches

$$\int_0^{\pi} f(2u) \cos u \, du,$$

by the definition of integrability, and the second sum has the same limit, so that the whole expression approaches zero. (Neither sum as it stands is exactly of the form that would ordinarily be written down in defining the definite integral,

but the discrepancy in each case amounts to half of a single term, and approaches zero in the limit.) As

$$\sin n t_r = (-1)^{r+1} \sin \frac{r\pi}{2n+1},$$

similar reasoning applies to the expression

$$[1/(2n+1)] \sum f(t_r) \sin n t_r.$$

If $m = 2n$, the summation going from $r = 0$ to $r = 2n-1$,

$$t_r = r\pi/n, \qquad \cos n t_r = \cos r\pi = (-1)^r,$$

$$\frac{1}{2n} \sum f(t_r) \cos n t_r$$

$$= \frac{1}{2\pi} \sum_{j=0}^{n-1} \frac{\pi}{n} f\left(\frac{2j\pi}{n}\right) - \frac{1}{2\pi} \sum_{j=1}^{n} \frac{\pi}{n} f\left(\frac{(2j-1)\pi}{n}\right).$$

Each sum in the last member approaches

$$\int_0^{\pi} f(2u)\, du,$$

and the whole expression approaches zero. As for the other expression in the statement of the lemma, $\sin n t_r = 0$ throughout.

Suppose $f(x)$ is a function of period 2π, bounded and integrable in the sense of Riemann, which vanishes for $x_0 - \eta \leq x \leq x_0 + \eta$. It will be seen that $S_n(x_0)$ *converges to the value* 0, whether the values of m entering into the definition of the sums $S_n(x)$ are even or odd. In both cases, $S_n(x)$ can be represented, artificially but with readily verifiable accuracy, by the single formula

$$S_n(x) = \frac{1}{m} \sum f(t_r) \sin n(t_r - x) \cot \frac{1}{2}(t_r - x)$$

$$+ \frac{1}{2m} [1 - (-1)^m] \sum f(t_r) \cos n(t_r - x).$$

As $\cos n(t_r - x) = \cos nx \cos n t_r + \sin nx \sin n t_r$, the last sum, taken with the factor $1/(2m)$, approaches zero uniformly for all values of x. And as $f(t) \cot \frac{1}{2}(t - x_0)$ is bounded

and integrable in the sense of Riemann as a function of t, the first factor vanishing where the second factor is large. application of Lemma II to this function shows that the first part of the expression for $S_n(x)$ approaches zero for $x = x_0$. If two functions, each bounded and integrable in the sense of Riemann, are identical from $x_0 - \eta$ to $x_0 + \eta$, the difference of their interpolating expressions converges toward zero at x_0; *if $f(x)$ is any function of period 2π, bounded and integrable in the sense of Riemann, the convergence of the corresponding interpolating sums $S_n(x)$ at any specified point, as m becomes infinite through odd or even values, depends only on the behavior of $f(x)$ in the neighborhood of the point in question.*

Now let $f(x)$ be of period 2π, bounded and integrable in the sense of Riemann, and identically zero for $\alpha - \eta < x < \beta + \eta$. and let attention be directed to the problem of showing that $S_n(x)$ converges toward zero uniformly for $\alpha \leqq x \leqq \beta$. It was noted in the last paragraph that $S_n(x)$ either is given identically by the expression

$$\frac{1}{m} \sum f(t_r) \sin n (t_r - x) \cot \frac{1}{2} (t_r - x),$$

or differs from it by an amount which approaches zero uniformly for all values of x as n becomes infinite. Furthermore, the terms resulting from the expansion of $\sin n(t_r - x)$ may be considered separately, and as the factors $\sin nx$ and $\cos nx$ are bounded, it is sufficient to demonstrate the uniform convergence of the expressions

$$\frac{1}{m} \sum f(t_r) \cos n t_r \cot \frac{1}{2} (t_r - x),$$

$$\frac{1}{m} \sum f(t_r) \sin n t_r \cot \frac{1}{2} (t_r - x),$$

as x in the argument of the cotangent ranges over the interval (α, β).

Suppose in the first place that $m = 2n + 1$. Let x be restricted to the interval $\alpha \leqq x \leqq \beta$. Let $C(x, t)$ be a function which vanishes when $|t - x|$ differs from an integral multiple of 2π by less than η, and is equal to $\cot \frac{1}{2}(t - x)$ elsewhere.

Then $f(t)\,C(x,\,t)$ is identical with $f(t)\cot\frac{1}{2}(t-x)$, under the restriction imposed on x, while $C(x,\,t)$ is bounded for all values of x and t, never exceeding $\cot\frac{1}{2}\eta$ in absolute value, and is R-integrable as a function of t. By adaptation of the formulas of an earlier paragraph,

$$[2\pi/(2n+1)]\sum f(t_r)\cos n\,t_r C(x,\,t_r)$$

is equal to

$$\sum_{j=0}^{n}\frac{2\pi}{2n+1}f\left(\frac{4j\pi}{2n+1}\right)\cos\frac{2j\pi}{2n+1}\,C\left(x,\frac{4j\pi}{2n+1}\right)$$

minus another expression of similar form. Let $u_j = 2j\pi/(2n+1)$, $j=0,1,\cdots,n$, and let $\Delta u_j = 2\pi/(2n+1)$ for $j=0,1,\cdots,n-1$, while $\Delta u_n = \pi/(2n+1)$. Then the sum just written down differs merely by the quantity

$$\frac{\pi}{2n+1}f\left(\frac{4n\pi}{2n+1}\right)\cos\frac{2n\pi}{2n+1}\,C\left(x,\;\frac{4n\pi}{2n+1}\right),$$

which approaches zero uniformly for all values of x, from

$$\sum_{j=0}^{n} f(2u_j)\cos u_j\,C(x,\,2u_j)\,\Delta u_j,$$

a sum corresponding to a subdivision of the interval $(0,\,\pi)$ into $n+1$ parts, not all of equal length, and defining the definite integral

$$\int_0^{\pi} f(2u)\cos u\,C(x,\,2u)\,du$$

in the limit.

The essential point for the present argument is that the sum differs from the integral by an amount which approaches zero uniformly with respect to x. Let M_j and m_j be the least upper bound and the greatest lower bound of the integrand over the jth sub-interval, P_j and p_j the corresponding bounds for the factor $f(2u)\cos u$, and Q_j and q_j those for the factor $C(x,\,2u)$. The difference between the sum and the integral does not exceed

$$\sum_{j=0}^{n}(M_j-m_j)\,\Delta u_j.$$

Let M be the least upper bound of $|f(2u)|$, as u varies without restriction, an upper bound consequently for $|f(2u)\cos u|$, and $Q(=\cot\frac{1}{2}\eta)$ the least upper bound of $|C(x, 2u)|$. Then

$$M_j - m_j \leq M(Q_j - q_j) + Q(P_j - p_j),$$

since for any u_1 and u_2 of the sub-interval in question

$$\begin{aligned} f(2u_2)\cos u_2\, C(x, 2u_2) - f(2u_1)\cos u_1\, C(x, 2u_1) \\ = f(2u_2)\cos u_2\,[C(x, 2u_2) - C(x, 2u_1)] \\ + C(x, 2u_1)\,[f(2u_2)\cos u_2 - f(2u_1)\cos u_1]. \end{aligned}$$

As $f(u)$ is by hypothesis integrable in the sense of Riemann, $f(2u)\cos u$ is R-integrable also. Let ϵ be an arbitrary positive quantity. By the property of integrability (more specifically, by Darboux's theorem) it is possible to choose n so large, and the intervals Δu_j in consequence uniformly so small, that

$$\sum_{j=0}^{n} (P_j - p_j)\,\Delta u_j < \frac{\epsilon}{2Q}, \quad \sum_{j=0}^{n} Q(P_j - p_j)\,\Delta u_j < \frac{\epsilon}{2}.$$

On the other hand,

$$\sum_{j=0}^{n} (Q_j - q_j)$$

can not exceed the total variation of $C(x, 2u)$ as u ranges over the interval $(0, \pi)$, a total variation which is finite *and independent of* x, being equal to $4Q$. Hence if n is taken so large that $2\pi/(2n+1) < \epsilon/(8QM)$, it follows that

$$\sum_{j=0}^{n} M(Q_j - q_j)\,\Delta u_j < \frac{\epsilon}{2}.$$

So there exists an N, independent of x, such that for $n \geq N$

$$\sum_{j=0}^{n} (M_j - m_j)\,\Delta u_j < \epsilon.$$

The other sum analogous to the one just treated is

$$\sum_{j=1}^{n} \frac{2\pi}{2n+1} f\left(\frac{2(2j-1)\pi}{2n+1}\right) \cos\frac{(2j-1)\pi}{2n+1}\, C\left(x, \frac{2(2j-1)\pi}{2n+1}\right).$$

By addition of the term

$$\frac{\pi}{2n+1} f\left(\frac{2(2n-1)\pi}{2n+1}\right) \cos\frac{(2n-1)\pi}{2n+1}\, C\left(x, \frac{2(2n-1)\pi}{2n+1}\right),$$

which approaches zero uniformly, this becomes a sum approximating the same definite integral as before, and corresponding to a subdivision of the interval of integration into $n-1$ parts each of length $2\pi/(2n+1)$ and one part of length $3\pi/(2n+1)$. The difference between the sum and the integral approaches zero uniformly, and the difference between the two sums does the same, which means that

$$\frac{1}{2n+1} \sum f(t_r) \cos n t_r \cot \frac{1}{2} (t_r - x)$$

converges uniformly toward zero. The corresponding expression with $\sin n t_r$ in place of $\cos n t_r$ can be treated in the same way. This completes the proof of uniform convergence of $S_n(x)$ for odd values of n.

For $m = 2n$, the discussion of the sum containing $\cos n t_r$ is slightly simplified in form by the fact that $\cos n t_r$ reduces to $(-1)^r$, and by the fact that $2\pi/m$ is an exact submultiple of π, but otherwise follows essentially the same lines as before. The sum with $\sin n t_r$ vanishes identically for all values of n, since $\sin n t_r = 0$.

The conclusion is embodied in

LEMMA III. *If $f(x)$ is a function of period 2π which is bounded and integrable in the sense of Riemann, and which vanishes identically for $\alpha - \eta \leq x \leq \beta + \eta$, the corresponding interpolating sums $S_n(x)$ converge uniformly toward zero for $\alpha \leq x \leq \beta$, as m becomes infinite through odd or even values, or both.*

On the formal side it may be pointed out in passing, and might have been noted before, that for m even $S_n(x)$ has the alternative representation

$$-\frac{1}{2n} \sin n x \sum f(t_r) \cos n t_r \cot \frac{1}{2} (t_r - x).$$

though the appearance of $\sin nx$ as a factor of the whole expression is illusory to the extent that corresponding to each value of x for which $\sin nx = 0$ there is a factor $\cot \frac{1}{2}(t_r - x)$ which becomes infinite.

Lemma III may be combined at once with Corollary IIa of Theorem I to give

THEOREM II. *If $f(x)$ is a function of period 2π, bounded and integrable in the sense of Riemann over a period, and continuous for $\alpha - \eta \leqq x < \beta + \eta$ with a modulus of continuity $\omega(\delta)$ such that* $\lim_{\delta=0} \omega(\delta) \log \delta = 0$, *the corresponding interpolating sums $S_n(x)$ converge uniformly toward $f(x)$ for $\alpha \leqq x \leqq \beta$, as m becomes infinite through odd or even values, or both.*

4. Convergence under hypothesis of limited variation

There are theorems of convergence for functions of limited variation, analogous to those obtained in the case of Fourier series. Let $f(x)$ be a function of period 2π, with limited variation over a period, and let x have a value, to be regarded for the time being as fixed, such that $f(t)$ is continuous for $t = x$. It will be shown that $S_n(x)$ converges toward $f(x)$ for the value of x specified.

Let $\varphi_1(t)$ and $\varphi_2(t)$ be the positive and negative variations of $f(t)$, measured from the point $t = -2\pi$, say, so that

$$f(t) = f(-2\pi) + \varphi_1(t) - \varphi_2(t),$$

while φ_1 and φ_2 are monotone increasing, and continuous for $t = x$. (The point x may of course be thought of without loss of generality as belonging to the interval $(0, 2\pi)$, and then there will be no occasion to take account of values of t less than -2π.) Consider the case $m = 2n + 1$. If x is one of the numbers t_r, for a specified value of n, $S_n(x) = f(x)$ exactly. Otherwise, let t_R be the number t_r which is nearest to x (or one of the two such numbers, if two are equally near), so that

$$t_R - \frac{\pi}{2n+1} \leqq x \leqq t_R + \frac{\pi}{2n+1}.$$

Let the summation over r in the formula representing $S_n(x)$ be extended from $R-n$ to $R+n$. The corresponding formula with $f(t_r)$ replaced by 1 represents 1 identically, since 1 is a trigonometric sum of order zero. If this identity is multiplied by the quantity $f(x)$, independent of r, and the result subtracted from the formula for $S_n(x)$, the difference $S_n(x)-f(x)$ is obtained in the form

$$S_n(x)-f(x)=\frac{1}{2n+1}\sum_{r=R-n}^{R+n}[f(t_r)-f(x)]\frac{\sin(n+\frac{1}{2})(t_r-x)}{\sin\frac{1}{2}(t_r-x)}.$$

The right-hand member is equal to the difference between

$$\frac{1}{2n+1}\sum_{r=R-n}^{R+n}[\varphi_1(t_r)-\varphi_1(x)]\frac{\sin(n+\frac{1}{2})(t_r-x)}{\sin\frac{1}{2}(t_r-x)}$$

and the similar expression with $\varphi_2(t_r)-\varphi_2(x)$ in place of $\varphi_1(t_r)-\varphi_1(x)$.

The quotient of sines never exceeds $2n+1$ in absolute value. In the term corresponding to the index $r=R$, t_R approaches x as n becomes infinite, and $\varphi_1(t_R)$ approaches $\varphi_1(x)$, by the continuity of φ_1, and the whole term, multiplied by $1/(2n+1)$, approaches zero and need not be further taken into account; it is sufficient to consider the sum from $R+1$ to $R+n$ and the corresponding sum from $R-n$ to $R-1$.

Since $t_{R+j}-x=t_R-x+[2j\pi/(2n+1)]$,

$$\sin\left(n+\frac{1}{2}\right)(t_{R+j}-x)=\sin\left[\left(n+\frac{1}{2}\right)(t_R-x)+j\pi\right]$$
$$=(-1)^j\sin\left(n+\frac{1}{2}\right)(t_R-x),$$

and is independent of j, except as to algebraic sign. If $\frac{1}{2}(t_{R+j}-x)$ is denoted by u_j, and if ϵ stands for 0 or 1 according as $\sin(n+\frac{1}{2})(t_R-x)$ is positive or negative,

$$\frac{1}{2n+1}\sum_{r=R+1}^{R+n}[\varphi_1(t_r)-\varphi_1(x)]\frac{\sin(n+\frac{1}{2})(t_r-x)}{\sin\frac{1}{2}(t_r-x)}$$
$$=(-1)^\epsilon\sum_{j=1}^{n}(-1)^j A_j B_j,$$

where

$$A_j = \varphi_1(x + 2u_j) - \varphi_1(x), \qquad B_j = \frac{|\sin(n+\frac{1}{2})(t_R - x)|}{(2n+1)\sin u_j}.$$

The numbers A_j, B_j are all positive or zero. As j increases, A_j increases or remains unchanged; u_j is always between 0 and $\pi/2$, $\sin u_j$ increases, and B_j decreases. Let V be the total variation of $f(x)$ over an interval of length 2π. Then the A's have the upper bound $\frac{1}{2}V$, since the positive and negative variations of a periodic function over a period are each equal to half the total variation; while B_j, which is equal to the absolute value of

$$\frac{\sin(n+\frac{1}{2})(t_{R+j} - x)}{(2n+1)\sin\frac{1}{2}(t_{R+j} - x)},$$

can not exceed unity.

After the analogy of the proof in the case of Fourier series, where the second law of the mean was employed, it would be natural here to use the method of summation by parts. The formulation can be slightly simplified, however, by reason of the fact that the sequences A_j, B_j are *both* monotone, though varying in opposite senses. In general, let $a_1, \cdots, a_p$, $b_1, \cdots, b_p$ be two sets of p numbers each, satisfying the conditions

$$a_1 \geqq a_2 \geqq \cdots \geqq a_p \geqq 0, \qquad b_1 \geqq b_2 \geqq \cdots \geqq b_p \geqq 0.$$

Let $c_j = b_1 - b_j$, so that

$$0 = c_1 \leqq c_2 \leqq \cdots \leqq c_p \leqq b_1.$$

Then

$$\begin{aligned} -\sum_{j=1}^{p}(-1)^j a_j b_{p-j+1} &= a_1 b_p - a_2 b_{p-1} + \cdots \\ &= a_1(b_1 - c_p) - a_2(b_1 - c_{p-1}) + \cdots \\ &= b_1(a_1 - a_2 + \cdots) - (a_1 c_p - a_2 c_{p-1} + \cdots). \end{aligned}$$

Since both the a's and the c's are non-negative and decrease monotonically, the value of each parenthesis in the last expression is positive or zero, and not greater than its leading term, in one case a_1, and in the other case $a_1 c_p$, which in turn is not greater than $a_1 b_1$. So the whole expression is the

difference of two non-negative quantities, neither of which can exceed $a_1 b_1$, and is itself not greater than $a_1 b_1$ in absolute value. To emphasize by words rather than by subscripts what is essential in the conclusion, *the absolute value of*

$$\sum_{j=1}^{p} (-1)^j a_j b_{p-j+1}$$

does not exceed the product of the largest of the a's by the largest of the b's.

Let ϵ be an arbitrary positive quantity. Corresponding to the continuity of φ_1, let $\delta > 0$ be chosen so that $|\varphi_1(t) - \varphi_1(x)| \leqq \frac{1}{2}\epsilon$ for $|t-x| < \delta$. Let the sum $\sum(-1)^j A_j B_j$ of the second paragraph preceding be broken up into two, $\sum'$ and $\sum''$, in the first of which j ranges over the values for which $2u_j < \delta$, while the remaining terms make up the second. In $\sum'$, the largest value of A_j is not greater than $\frac{1}{2}\epsilon$, while B_j never exceeds 1, so that by the conclusion of the last paragraph, with suitable adaptation of the subscripts, $|\sum'| \leqq \frac{1}{2}\epsilon$. In $\sum''$, $A_j \leqq \frac{1}{2} V$ throughout, while $B_j \leqq 1/[(2n+1)\sin\frac{1}{2}\delta]$: consequently

$$|\textstyle\sum''| < V \Big/ \left[2(2n+1)\sin\frac{1}{2}\delta\right],$$

which is less than $\frac{1}{2}\epsilon$ as soon as n is sufficiently large. Hence $\sum' + \sum''$, or, in the earlier notation, the sum from $R+1$ to $R+n$, multiplied by $1/(2n+1)$, approaches zero as n becomes infinite. The sum from $R-n$ to $R-1$ and the corresponding sums with φ_2 in place of φ_1 can be treated in the same way, to show that $S_n(x) - f(x)$ converges toward zero.

If $f(x)$ is continuous everywhere, the uniform continuity of φ_1 and φ_2, bearing on the choice of δ in the preceding paragraph, and, as a detail, on the convergence toward zero of the single term in each sum for which $r = R$, yields at once uniform convergence of $S_n(x)$ toward $f(x)$.

Finally, a precisely similar argument can be carried through for the case of an even number of interpolating points, $m = 2n$.

The conclusion is

THEOREM III. *If $f(x)$ is a function of period 2π having limited variation over a period, $S_n(x)$ converges toward $f(x)$ at every point where $f(x)$ is continuous, when m becomes infinite through odd or even values, or both; if $f(x)$, still supposed to be of limited variation, is continuous everywhere, the convergence is uniform for all values of x.*

This result can be generalized immediately by reference to Lemma III, and the more elementary fact that convergence at any point depends only on the behavior of the function in the neighborhood of the point:

COROLLARY. *If $f(x)$ is a function of period 2π, bounded and integrable in the sense of Riemann over a period, of limited variation for $x_0 - \eta \leq x \leq x_0 + \eta$, and continuous for $x = x_0$, $S_n(x_0)$ converges toward $f(x_0)$; if $f(x)$ is continuous and of limited variation for $\alpha - \eta \leq x \leq \beta + \eta$, the hypothesis remaining otherwise unchanged, $S_n(x)$ converges toward $f(x)$ uniformly for $\alpha \leq x \leq \beta$.*

5. Degree of convergence under hypotheses involving limited variation

Let $f(x)$ once more be of limited variation over a period, with total variation V. Let $\varphi_1(x)$ be the positive variation of $f(x)$ from 0 to x when $x > 0$, and minus the positive variation from x to 0 when $x < 0$, and let $\varphi_2(x)$ be the correspondingly defined negative variation function, so that

$$f(x) = f(0) + \varphi_1(x) - \varphi_2(x), \qquad \varphi_1(0) = \varphi_2(0) = 0.$$

Let $m = 2n+1$, $n > 0$. Consider the sum $\sum f(t_r) \cos n t_r$, extended specifically from $r = -n$ to $r = n$. The sum $\sum f(0) \cos n t_r$ vanishes. In $\sum \varphi_1(t_r) \cos n t_r$, the term corresponding to $r = 0$ vanishes with $\varphi_1(0)$. It has already been observed that $\cos n t_r = (-1)^r \cos [r\pi/(2n+1)]$. In the sum

$$\sum_{r=1}^{n} \varphi_1(t_r) \cos n t_r,$$

the factor $\varphi_1(t_r)$ increases monotonically, having the upper bound $\varphi_1(\pi)$, while, as $r\pi/(2n+1)$ remains within the limits $(0, \pi/2)$, $|\cos n t_r|$ decreases monotonically, having the upper bound 1. Consequently, by the reasoning of an earlier paragraph,

$$\left|\sum_{r=1}^{n} \varphi_1(t_r) \cos n t_r\right| \leqq \varphi_1(\pi).$$

Similarly, the absolute value of the sum from $-n$ to -1 does not exceed $|\varphi_1(-\pi)|$. So the whole sum from $-n$ to $+n$ does not exceed

$$\varphi_1(\pi) + |\varphi_1(-\pi)|, \qquad \varphi_1(\pi) - \varphi_1(-\pi) = \frac{1}{2} V.$$

After analogous reasoning with φ_2 in place of φ_1, it is concluded that

$$\left|\sum f(t_r) \cos n t_r\right| \leqq V.$$

There is a corresponding inequality

$$\left|\sum f(t_r) \sin n t_r\right| \leqq V,$$

the proof being simplified in this case by the fact that the monotone sequences $\varphi_1(t_r)$, $|\varphi_2(t_r)|$, $|\sin n t_r|$ vary in the same sense.

For $m = 2n$ a still simpler calculation, based on the fact that $\cos n t_r = (-1)^r$, shows that $\left|\sum f(t_r) \cos n t_r\right| \leqq V$, while $\sum f(t_r) \sin n t_r = 0$.

Disregarding the special simplicity of the last observation, it is possible to state comprehensively

LEMMA IV. *If $f(x)$ is a function of period 2π, with limited variation over a period, its total variation being V, then*

$$\left|\sum f(t_r) \cos n t_r\right| \leqq V, \qquad \left|\sum f(t_r) \sin n t_r\right| \leqq V,$$

whether m is odd $(m = 2n+1)$ or even $(m = 2n)$.

Let $f(x)$ be of limited variation over a period, with total variation V, and identically zero for $\alpha - \eta \leqq x \leqq \beta + \eta$.

It has already been pointed out in substance that if e_m is a symbol denoting 1 or 0 according as m is odd or even, $S_n(x)$ satisfies the identity

$$m S_n(x) = \cos nx \sum f(t_r) \cot \tfrac{1}{2}(t_r - x) \sin n t_r$$
$$- \sin nx \sum f(t_r) \cot \tfrac{1}{2}(t_r - x) \cos n t_r$$
$$+ e_m \cos nx \sum f(t_r) \cos n t_r + e_m \sin nx \sum f(t_r) \sin n t_r.$$

For any value of x in (α, β), the product $f(t) \cot \frac{1}{2}(t - x)$ is of limited variation, regarded as a function of t, and its total variation does not exceed the product of V by a quantity depending only on η. So the application of Lemma IV to each of the four sums in turn leads to the

Corollary. *If $f(x)$ is a function of period 2π with limited variation, the total variation over a period being V, and if $f(x)$ vanishes identically for $\alpha - \eta \leq x \leq \beta + \eta$, then*

$$|S_n(x)| \leq C_\eta V / n$$

for $\alpha \leq x \leq \beta$, where C_η depends only on η.

This Corollary in turn may be associated with Corollary II of Theorem I, in the light of the discussion leading up to Theorem Va in Chapter II, to yield

Theorem IV. *If the function $f(x)$, of period 2π, is continuous with modulus of continuity $\omega(\delta)$ for $\alpha - \eta \leq x \leq \beta + \eta$, where $\omega(\delta) \to 0$ for $\delta \to 0$, and of limited variation over the rest of a period, then*

$$|f(x) - S_n(x)| \leq c\,\omega(2\pi/n) \log n$$

for $\alpha \leq x \leq \beta$, if n is large enough so that $\omega(2\pi/n)$ has a meaning, c being a constant which depends neither on x nor on n.

Lemma IV is analogous to Theorem III of Chapter II, the sums in the present lemma, multiplied by a quantity of the order of $1/n$ to represent the length of the interval between successive points t_r, corresponding to the integrals in the

earlier formulation, though as the length of the interval is either exactly or approximately π/n, according as m is even or odd, the agreement does not extend to the numerical values obtained for the constants in the right-hand members of the inequalities. It is sufficiently noteworthy, however, that the agreement should be as close at it is, since for any particular value of n the number of points used is not large enough to make it all evident *a priori* that the sums give close approximations to the values of the integrals, in view of the presence of the factors $\cos nx$, $\sin nx$ under the sign of integration.

The analogy can be extended to the conclusions of Corollary I of the theorem referred to. The proof is perhaps most readily given by means of the Corollary itself. *Suppose $f(x)$ has a pth derivative with limited variation.* Then $f(x)$ is represented by a convergent Fourier series; if the Fourier coefficients are denoted by α_k, β_k,

$$f(x) = \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} (\alpha_k \cos kx + \beta_k \sin kx).$$

The series may be used to represent $f(t_r)$ in evaluating the sums $\sum f(t_r) \cos nt_r$, $\sum f(t_r) \sin nt_r$. Let $m = 2n+1$. The expression $\sum \cos kt_r \cos nt_r$ vanishes unless $k-n$ or $k+n$ is an integral multiple of $2n+1$, that is, unless k has one of the values n, $n+1$, $3n+1$, $3n+2$, $5n+2$, $5n+3, \cdots$. The value of the sum in each of these cases is $(2n+1)/2$. The sum $\sum \sin kt_r \cos nt_r$ is zero in all cases. Consequently

$$\sum f(t_r) \cos nt_r = [(2n+1)/2]\,(\alpha_n + \alpha_{n+1} + \alpha_{3n+1} + \alpha_{3n+2} + \cdots).$$

Similarly,

$$\sum f(t_r) \sin nt_r = [(2n+1)/2]\,(\beta_n - \beta_{n+1} + \beta_{3n+1} - \beta_{3n+2} + \cdots).$$

But by the Corollary cited,

$$|\alpha_k| \leq V/(\pi k^{p+1}), \qquad |\beta_k| \leq V/(\pi k^{p+1}),$$

if V is the total variation of $f^{(p)}(x)$. So

$$|\alpha_n + \alpha_{n+1} + \alpha_{3n+1} + \alpha_{3n+2} \cdots|$$
$$\leqq \frac{V}{\pi}\left[\frac{1}{n^{p+1}} + \frac{1}{(n+1)^{p+1}} + \frac{1}{(3n+1)^{p+1}} + \frac{1}{(3n+2)^{p+1}} + \cdots\right]$$
$$< \frac{V}{\pi}\left[\frac{1}{n^{p+1}} + \frac{1}{n^{p+1}} + \frac{1}{(3n)^{p+1}} + \frac{1}{(3n)^{p+1}} + \cdots\right]$$
$$= \frac{2V}{\pi n^{p+1}}\left[1 + \frac{1}{3^{p+1}} + \frac{1}{5^{p+1}} + \cdots\right],$$

the series in the last pair of brackets being convergent for $p \geqq 1$, and having incidentally, for $p > 1$, the upper bound $1 + 1/3^2 + 1/5^2 + \cdots$, independent of p. A similar inequality holds for the series of β's. If $m = 2n$.

$$\sum f(t_r) \cos nt_r = 2n(\alpha_n + \alpha_{3n} + \alpha_{5n} + \cdots),$$

and $\sum f(t_r) \sin nt_r = 0$. *Whether m is odd or even*, therefore,

$$|\sum f(t_r) \cos nt_r| < CV/n^p, \qquad |\sum f(t_r) \sin nt_r| < CV/n^p,$$

where C is an absolute constant.

This result, supplemented by a type of argument which has been used a number of times already, gives at once certain information with regard to the degree of convergence of interpolating sums $S_n(x)$. For example, if $f(x)$ has a first derivative which is of limited variation over a period, and if $f(x)$ vanishes identically for $\alpha - \eta \leqq x \leqq \beta + \eta$, then $S_n(x)$ can not exceed a constant multiple of $1/n^2$ for $\alpha \leqq x \leqq \beta$; if $f(x)$ has a first derivative which is of limited variation over a period, and if $f'(x)$ is continuous with modulus of continuity $\omega(\delta)$ for $\alpha - \eta \leqq x \leqq \beta + \eta$, but not identically constant over the interval, then $|f(x) - S_n(x)|$ can not exceed a constant multiple of $(1/n)\,\omega(2\pi/n) \log n$ for $\alpha \leqq x \leqq \beta$. But such observations are of secondary interest; and it is *not* possible to pass over immediately to a proposition analogous to Theorem IV of Chapter II, for the transition from a specified interpolating sum $S_n(x)$ to one of higher order does not consist merely in the inclusion of additional terms directly subject

to the inequalities that have been under discussion. For the problem thus suggested a different procedure is required. The detailed treatment will be limited to the next case in order, that of a function having a first derivative of limited variation over a period (or expressible as the integral of a function of limited variation), and not otherwise restricted.

The key to the discussion of this case is the fact that if $f(x)$ is the integral of a function of limited variation over an interval $a \leqq x \leqq b$, the quotient $[f(x)-f(a)]/(x-a)$ (defined in any way, by the value $f'(a+)$ or otherwise, for $x=a$) is likewise of limited variation over the interval. Let

$$f(x) = f(a) + \int_a^x \varphi(x)\,dx,$$

where $\varphi(x) = \varphi_1(x) - \varphi_2(x)$, and φ_1 and φ_2 are bounded non-decreasing functions from a to b. Let

$$\Phi_1(x) = \frac{1}{x-a}\int_a^x \varphi_1(x)\,dx$$

for $a < x \leqq b$, while $\Phi_1(a) = \varphi_1(a)$, and let a function $\Phi_2(x)$ be similarly defined in terms of $\varphi_2(x)$. Then, for $x > a$,

$$\frac{f(x)-f(a)}{x-a} = \Phi_1(x) - \Phi_2(x).$$

The assertion with regard to the quotient on the left will be proved if it is shown that $\Phi_1(x)$ and $\Phi_2(x)$ are non-decreasing; the fact is rather obvious, and in formulating a proof it is clearly sufficient to consider one of the two functions. For any value of $x > a$,

$$(x-a)\,\varphi_1(a) \leqq \int_a^x \varphi_1(x)\,dx \leqq (x-a)\,\varphi_1(x),$$

by reason of the monotone character of φ_1, so that $\Phi_1(x) \geqq \varphi_1(a) = \Phi_1(a)$. Let x_1 and x_2 be any two values such that $a < x_1 < x_2 \leqq b$. Then

$$\int_a^{x_1} \varphi_1(x)\,dx = (x_1-a)\,X_1,$$

$$\int_a^{x_2} \varphi_1(x)\,dx = \int_a^{x_1} + \int_{x_1}^{x_2} = (x_1-a)\,X_1 + (x_2-x_1)\,X_2,$$

where

$$\varphi_1(a) < X_1 \leqq \varphi_1(x_1) \leqq X_2 \leqq \varphi_1(x_2).$$

So

$$\int_a^{x_2} \varphi_1(x)\,dx > (x_1 - a) X_1 + (x_2 - x_1) X_1 = (x_2 - a) X_1.$$

whence $\Phi_1(x_2) \geqq X_1$; as $X_1 = \Phi_1(x_1)$, the desired conclusion is established. Similarly, $[f(x) - f(b)]/(b - x)$ is of limited variation.

Now let $f(x)$ be of period 2π, and (for simplicity of statement) provided with a first derivative of limited variation over a period. The function $(x-a)\cot\frac{1}{2}(x-a)$, defined so as to be continuous for $x = a$, has a continuous derivative for $a \leqq x \leqq a+\pi$, and so is of limited variation over this interval. Hence the product

$$\{[f(x)-f(a)]/(x-a)\}\left\{(x-a)\cot\frac{1}{2}(x-a)\right\}$$
$$= [f(x)-f(a)]\cot\frac{1}{2}(x-a)$$

is of limited variation over $(a, a+\pi)$, since this is true of each of the expressions in braces. Similarly, $(x-a-2\pi)\cot\frac{1}{2}(x-a)$ is of limited variation for $a+\pi < x \leqq a+2\pi$; as $f(a+2\pi) = f(a)$, the expression $[f(x)-f(a)]/(a+2\pi-x)$ is the same as $[f(x)-f(a+2\pi)]/(a+2\pi-x)$, which is of limited variation over $(a+\pi, a+2\pi)$; and so $[f(x)-f(a)]\cot\frac{1}{2}(x-a)$ is of limited variation over this interval also. In summary, and in slightly different notation, the expression $[f(t) - f(x)]\cot\frac{1}{2}(t - x)$, regarded as a function of t, is of limited variation over any interval of length 2π.

Let the interpolating sum for $f(x)$ be expressed once more in the form used in connection with the Corollary of Lemma IV. The interpolating sum for a constant reproduces the constant identically; since $f(x)$ is a constant with respect to the index of summation r, the quantity given by the formula reduces identically to $mf(x)$, if $f(t_r)$ is replaced by $f(x)$ under the sign of summation, and

$$
\begin{aligned}
m[S_n(x)-f(x)] \\
&= \cos nx \sum [f(t_r)-f(x)] \cot \tfrac{1}{2}(t_r-x) \sin n t_r \\
&\quad - \sin nx \sum [f(t_r)-f(x)] \cot \tfrac{1}{2}(t_r-x) \cos n t_r \\
&\quad + e_m \cos nx \sum [f(t_r)-f(x)] \cos n t_r \\
&\quad + e_m \sin nx \sum [f(t_r)-f(x)] \sin n t_r .
\end{aligned}
$$

But $f(t)-f(x)$, as well as $[f(t)-f(x)]\cot\frac{1}{2}(t-x)$, is ot limited variation with respect to t, and consequently Lemma IV is applicable to each of the sums on the right. Closer inspection shows that the total variation in each case is not greater than a constant multiple of the total variation of $f'(x)$ over a period. If this fact is incorporated in the statement of the result, the conclusion may be formulated as

THEOREM V. *If $f(x)$ is a function of period 2π having a first derivative with limited variation, the total variation of $f'(x)$ over a period being V, then*

$$|f(x)-S_n(x)| \leqq CV/n,$$

where C is an absolute constant.

The corresponding analysis for the case of functions having higher derivatives will not be carried through here. It may be pointed out, however, as a first stage in the extension, that if $f(x)$ has a continuous non-decreasing second derivative for $a \leqq x \leqq b$, then $[f(x)-f(a)]/(x-a)$, considered to have the value $f'(a)$ for $x=a$, has a non-decreasing first derivative over the same interval. For $x=a$,

$$\frac{d}{dx}\left[\frac{f(x)-f(a)}{x-a}\right] = \frac{1}{2} f''(a),$$

as may be seen by calculating the derivative from first principles and applying the extended mean value theorem in the process. For $x>a$,

$$
\begin{aligned}
\frac{d}{dx}\left[\frac{f(x)-f(a)}{x-a}\right] &= \frac{(x-a)f'(x)-[f(x)-f(a)]}{(x-a)^2} \\
&= \frac{1}{2} f''(\xi) \geqq \frac{1}{2} f''(a),
\end{aligned}
$$

the number ξ lying between a and x. For $x > a$, furthermore,

$$\frac{d^2}{dx^2}\left[\frac{f(x)-f(a)}{x-a}\right] = \frac{(x-a)^2 f''(x) - 2(x-a)f'(x) + 2[f(x)-f(a)]}{(x-a)^3}$$
$$= \frac{f''(x) - f''(\xi)}{x-a} > 0.$$

The monotone character of the first derivative thus becomes apparent. It follows that if $f(x)$ has a continuous second derivative of limited variation, $[f(x)-f(a)]/(x-a)$ has a first derivative of limited variation, and connection can be made with the facts previously ascertained as to functions satisfying the latter condition.

6. Formula of interpolation analogous to the Fejér mean

A considerable part of Chapter II was devoted to a discussion of the arithmetic mean of the partial sums of the Fourier series. There is a corresponding formula in the case of interpolation, possessing many analogous properties, with the outstanding exception that it is *not* an arithmetic mean of a sequence of the interpolating sums previously studied. It is to be defined and examined on its own merits, with only incidental reference to the content of the earlier part of the present chapter.

Let n be an arbitrary positive integer, and let $t_r = 2r\pi/n$, for any integral value of r. In comparison with the earlier notation, m is now to be taken equal to n, instead of $2n+1$ or $2n$; the sign $\sum$ will be understood to refer to summation over n successive values of the index r, when there is no indication to the contrary; a separate symbol m is no longer needed; and there is no occasion to distinguish between odd and even values of n.

Let $f(x)$ be an arbitrary function of period 2π. The interpolating formula in question is

$$\sigma_n(x) = \sum f(t_r)\,\frac{\sin^2 \frac{1}{2} n(t_r - x)}{n^2 \sin^2 \frac{1}{2}(t_r - x)},$$

with the understanding that each term is defined so as to be continuous wherever its denominator vanishes. For $x = t_q$, one term of the sum reduces to $f(t_q)$, namely that in which $r = q$, or in which r differs from q by an integral multiple of n, while each of the other terms becomes zero, and consequently

$$\sigma_n(t_q) = f(t_q).$$

So the interpolating property is apparent at the outset.

It is an almost immediate consequence of identities previously employed that

$$\left(\sin^2 \frac{1}{2} n v\right) \Big/ \left(\sin^2 \frac{1}{2} v\right)$$
$$= 2\left[\frac{1}{2} n + (n-1)\cos v + (n-2)\cos 2v + \cdots + \cos(n-1)v\right].$$

Hence, as a result of the substitution $v = t_r - x$,

$$\sigma_n(x) = \frac{1}{2} a_0 + a_1 \cos x + a_2 \cos 2x + \cdots + a_{n-1}\cos(n-1)x$$
$$+ b_1 \sin x + b_2 \sin 2x + \cdots + b_{n-1} \sin(n-1)x,$$

with the coefficients

$$a_k = \frac{2(n-k)}{n^2} \sum f(t_r)\cos k t_r,$$

$$b_k = \frac{2(n-k)}{n^2} \sum f(t_r)\sin k t_r.$$

The expression is a trigonometric sum, but of order about twice as high as would be required merely for the purpose of obtaining coincidence at n points.

If $f(x) = 1$, all the coefficients reduce to zero except a_0, while $\frac{1}{2} a_0 = 1$, so that in this case $\sigma_n(x) = 1$, or, if $(\sin^2 n u)/(\sin^2 u)$ is denoted by $\Phi_n(u)$,

$$\sum \frac{\sin^2 \frac{1}{2} n (t_r - x)}{n^2 \sin^2 \frac{1}{2}(t_r - x)} = \sum (1/n^2)\, \Phi_n\left[\frac{1}{2}(t_r - x)\right] = 1.$$

Hence it follows further, inasmuch as $\Phi_n(u)$ is never negative, that if $f(x)$ is any function having M as an upper bound for its absolute value,

$$|\sigma_n(x)| = \left|\sum (1/n^2) f(t_r)\, \Phi_n\left[\frac{1}{2}(t_r - x)\right]\right|$$
$$\leqq \sum (M/n^2)\, \Phi_n\left[\frac{1}{2}(t_r - x)\right] = M.$$

Like the arithmetic mean associated with the Fourier series, the present $\sigma_n(x)$ converges uniformly toward $f(x)$, as n becomes infinite, if $f(x)$ is everywhere continuous, and converges at points of continuity under more general hypotheses as to the behavior of the function elsewhere. Inasmuch as the sums depend on the values of the function at isolated points, however, it is necessary to impose some restriction on the values which it may take on point by point, not merely to require that it be summable, or in other words that it have a finite *mean* value.

Let $f(t)$ be continuous for $t = x$, and let $|f(t)| < M$ everywhere. By a device already used on a number of occasions, the error of $\sigma_n(x)$ can be expressed in the form

$$\sigma_n(x) - f(x) = \sum (1/n^2)\, [f(t_r) - f(x)]\, \Phi_n\left[\frac{1}{2}(t_r - x)\right].$$

Let ϵ be an arbitrary positive quantity, and let δ be a positive number such that $|f(t) - f(x)| < \frac{1}{2}\epsilon$ for $|t - x| < \delta$. Let the sum in the right-hand member of the identity for $\sigma_n(x) - f(x)$ be written in the form $\sum' + \sum''$, where $\sum'$ denotes summation over those values of r for which $t_r - x$ differs from an integral multiple of 2π by less than δ, and $\sum''$ stands for a summation covering the remaining values of r. In $\sum'$, $|f(t_r) - f(x)| < \frac{1}{2}\epsilon$, so that

$$\left|\sum' (1/n^2)[f(t_r) - f(x)]\, \Phi_n\left[\frac{1}{2}(t_r - x)\right]\right|$$
$$< \sum' (1/n^2)(\epsilon/2)\, \Phi_n\left[\frac{1}{2}(t_r - x)\right]$$
$$\leqq \sum (1/n^2)(\epsilon/2)\, \Phi_n\left[\frac{1}{2}(t_r - x)\right] = \frac{1}{2}\epsilon.$$

In $\sum''$, $\Phi_n[\frac{1}{2}(t_r - x)] \leqq 1/(\sin^2 \frac{1}{2}\delta)$, and $|f(t_r) - f(x)| \leqq 2M$, while the number of terms can not exceed n, and consequently

$$\left|\sum''(1/n^2)[f(t_r)-f(x)]\,\Phi_n\left[\frac{1}{2}(t_r-x)\right]\right| \leq 2M/\left(n\sin^2\frac{1}{2}\delta\right),$$

which is less than $\frac{1}{2}\epsilon$ as soon as n is sufficiently large. This establishes the fact of convergence. It is sufficient that n surpass a bound depending only on M and δ; if $f(x)$ is continuous everywhere, δ can be chosen independently of x, and the convergence is uniform. If $f(x)$ is continuous for $\alpha-\eta \leq x \leq \beta+\eta$, without being necessarily continuous everywhere, it is possible to choose a $\delta \leq \eta$ which shall be valid for all values of x in the interval $\alpha \leq x \leq \beta$, and the convergence is uniform over the latter interval. Finally, if $f(x)$, instead of being merely continuous for $\alpha-\eta \leq x \leq \beta+\eta$, is identically zero there, and if x is given a value belonging to the interval (α, β), $\Phi_n[\frac{1}{2}(t_r-x)] \leq 1/(\sin^2\frac{1}{2}\eta)$ in all terms in which $f(t_r) \neq 0$, and

$$|\sigma_n(x)| \leq M/(n\sin^2\tfrac{1}{2}\eta).$$

The results may be summarized in

THEOREM VI. *If $f(x)$ is a bounded function of period 2π, $\sigma_n(x)$ converges toward $f(x)$ at every point at which $f(x)$ is continuous. If $f(x)$ is continuous everywhere, the convergence is uniform everywhere. If $f(x)$ is continuous for $\alpha-\eta \leq x \leq \beta+\eta$, the convergence is uniform for $\alpha \leq x \leq \beta$. If $f(x)$ is identically zero for $\alpha-\eta \leq x \leq \beta+\eta$, $|\sigma_n(x)| \leq C_\eta M/n$ for $\alpha \leq x \leq \beta$, where M is an upper bound for $|f(x)|$, and C_η is a constant depending only on η.*

Further discussion of the degree of convergence of $\sigma_n(x)$ will be limited to the case in which $f(x)$ satisfies the condition

$$|f(x_2)-f(x_1)| \leq \lambda|x_2-x_1|.$$

Suppose first that this condition is satisfied everywhere. Then

$$|\sigma_n(x)-f(x)| \leq (\lambda/n^2)\sum'|t_r-x|\,\Phi_n\left[\frac{1}{2}(t_r-x)\right].$$

For any particular value of x, let t_R be that one of the numbers t_r which is nearest to x, or one of the two nearest, if x is equally distant from two of them, so that $t_R-(\pi/n) \leq x \leq t_R+(\pi/n)$. Let the summation be thought of as

extended specifically over the n values of r for which $x-\pi \leq t_r < x+\pi$. The precise expressions for the extreme values of r in terms of R and n will vary according to circumstances; they will be approximately $R \pm \frac{1}{2} n$, and it is sufficient for the purposes of the present argument to note that they will certainly be between $R-n$ and $R+n$, for any $n > 1$.

For any value of v,

$$\Phi_n\left(\frac{1}{2} v\right) = 2\left[\frac{1}{2} n+(n-1)\cos v+(n-2)\cos 2v+\cdots+\cos(n-1)v\right]$$
$$\leq 2\left[\frac{1}{2} n+(n-1)+(n-2)+\cdots+1\right] = n^2,$$

and hence

$$\left|\frac{\sin \frac{1}{2} n v}{\sin \frac{1}{2} v}\right| = \left[\Phi_n\left(\frac{1}{2} v\right)\right]^{1/2} \leq n.$$

Furthermore, $0 \leq v/\sin\frac{1}{2}v \leq \pi$ for $v \leq \pi$. Throughout this interval, consequently

$$|v \Phi_n\left(\frac{1}{2} v\right)| = \frac{v}{\sin \frac{1}{2} v} \cdot \left|\frac{\sin \frac{1}{2} n v}{\sin \frac{1}{2} v}\right| \cdot \left|\sin \frac{1}{2} n v\right| \leq \pi \cdot n \cdot 1,$$

and at the same time (for $v \neq 0$)

$$|v \Phi_n\left(\frac{1}{2} v\right)| = \frac{1}{v} \cdot \frac{v^2}{\sin^2 \frac{1}{2} v} \cdot \sin^2 \frac{1}{2} n v \leq \frac{1}{v} \cdot \pi^2 \cdot 1.$$

These relations are to be used in connection with the inequality for $\sigma_n(x)-f(x)$ in the second paragraph preceding. If x is one of the numbers t_r, $\sigma_n(x)=f(x)$, and there is no further question as to the magnitude of the error. If x does not coincide with a t_r, there are just two values of r in the summation for which $|t_r-x| < 2\pi/n$. In the terms of the sum corresponding to these values of r it is sufficient to apply the relation $|v \Phi_n(\frac{1}{2} v)| \leq n\pi$, or $|t_r-x| \Phi_n[\frac{1}{2}(t_r-x)] \leq n\pi$. For the other points t_r, in the order of increasing distance from x on each side, the values of $|t_r-x|$ are successively greater than the numbers $2\pi/n$, $4\pi/n$, $6\pi/n$, $\cdots$, and by the relation $|v \Phi_n(\frac{1}{2} v)| \leq \pi^2/|v|$ the corresponding values of $|t_r-x| \Phi_n[\frac{1}{2}(t_r-x)]$ are respectively less than $n\pi/2$, $n\pi/4$, $n\pi/6$, $\cdots$. Hence, as the sum involves not more than n points t_r on each side of x,

$$\sum |t_r - x| \, \Phi_n\left[\frac{1}{2}(t_r - x)\right]$$
$$< 2n\pi + 2[n\pi/2 + n\pi/4 + n\pi/6 + \cdots + n\pi/(2n)]$$
$$= n\pi\left[2 + 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right]$$
$$\leq n\pi\left[3 + \int_1^n \frac{du}{u}\right] = n\pi(3 + \log n),$$

which for $n \geq 2$ does not exceed a constant multiple of $n \log n$. This makes it possible to state

THEOREM VII. *If $f(x)$ is a function of period 2π satisfying everywhere the condition*

$$|f(x_2) - f(x_1)| \leq \lambda |x_2 - x_1|,$$

then, for all values of $n \geq 2$,

$$|f(x) - \sigma_n(x)| < (C\lambda \log n)/n,$$

where C is an absolute constant.

The conclusion may be generalized at once by combination with the last assertion of Theorem VI, to yield the

COROLLARY. *If $f(x)$ is a bounded function of period 2π satisfying the condition*

$$|f(x_2) - f(x_1)| \leq \lambda |x_2 - x_1|$$

throughout the interval $\alpha - \eta \leq x \leq \beta + \eta$, then

$$|f(x) - \sigma_n(x)| < (c\lambda \log n)/n$$

for $\alpha \leq x \leq \beta$, where c is a constant depending neither on x nor on n.

In conclusion, it may be pointed out as a peculiarity of the interpolating sum $\sigma_n(x)$ that *its derivative vanishes at each of the points t_q* (See L. Fejér, Göttinger Nachrichten (1916), pp. 66–91, especially pp. 87–91). For in the identity

$$\sigma_n(x) = \sum (1/n^2) f(t_r) \, \Phi_n\left[\frac{1}{2}(t_r - x)\right]$$

it is evident from the expression of Φ_n in fractional form, or can be verified by differentiating this expression, that

$\Phi_n[\frac{1}{2}(t_r - x)]$ has a double root for $x = t_q$ if t_r is not congruent to t_q modulo 2π, while the representation of $\Phi_n[\frac{1}{2}(t_q - x)]$ as a sum of cosines shows that it also has a vanishing derivative for $x = t_q$.

7. Polynomial interpolation

As was indicated in the opening paragraph of the chapter, the methods that have been set forth are not adapted to the study of the problem of polynomial interpolation with equally spaced points. That problem is analogous rather to the theory of Taylor's series, whether treated by means of Taylor's theorem with the remainder for real variables, or by Cauchy's theorem in the complex plane. A simple change of variable, however, serves to carry over the formulas of trigonometric interpolation to a case of polynomial interpolation with *unequally* spaced points distributed in a certain way. Suppose namely that a function $f(x)$ is defined for $-1 \leq x \leq 1$. Then $f(\cos\theta)$ is a function defined for all values of θ. It is an even function of θ, and inspection of the formulas defining the coefficients in the interpolating sums $S_n(\theta)$ and $\sigma_n(\theta)$ shows at once that these sums involve only cosines, and so may be regarded as polynomials in $\cos\theta$. If $x = \cos\theta$, $S_n(\theta)$ and $\sigma_n(\theta)$ may be denoted by $P_n(x)$ and $\pi_n(x)$ respectively. They are polynomials agreeing in value with $f(x)$ for a set of values of x corresponding to equally spaced values of θ. The polynomial $\pi_n(x)$, unlike the interpolating polynomial of minimum degree for equally spaced values of x, converges in the case of every continuous function $f(x)$; its degree however is approximately twice as great as the number of points for which coincidence is obtained. The expression $P_n(x)$, on the other hand, is an interpolating polynomial of minimum degree, and while it is not convergent for every continuous function, it converges far more generally than the corresponding polynomial with equally spaced points. It is unnecessary to enumerate the further theorems on convergence and degree of convergence which would be obtained by following out the transformation of variable in detail.

CHAPTER V

INTRODUCTION TO THE GEOMETRY OF FUNCTION SPACE

1. The notions of distance and orthogonality

In Chapter III, attention was directed to the problem of the approximate representation of a given function by means of linear combinations of other given functions, according to the criterion of least squares. If $f(x)$ is a given function over an interval (a, b), and if $p_1(x), p_2(x), \cdots, p_m(x)$ are m linearly independent functions over the same interval, the coefficients $c_1, \cdots, c_m$ in an expression

$$\varphi(x) = c_1 p_1(x) + c_2 p_2(x) + \cdots + c_m p_m(x)$$

are to be determined so that

$$\int_a^b [f(x) - \varphi(x)]^2 \, dx$$

shall be a minimum. The value of the integral is taken as a measure of the discrepancy between the functions $f(x)$ and $\varphi(x)$.

The problem is the same in principle as that of the method of least squares for the approximate solution of a set of linear equations. Suppose there are n equations in m unknowns, $n > m$:

$$\begin{aligned} a_{11} x_1 + a_{21} x_2 + \cdots + a_{m1} x_m &= b_1, \\ a_{12} x_1 + a_{22} x_2 + \cdots + a_{m2} x_m &= b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{1n} x_1 + a_{2n} x_2 + \cdots + a_{mn} x_m &= b_n. \end{aligned}$$

The set of numbers $(b_1, b_2, \cdots, b_n)$ may be regarded as constituting a function $b(k) = b_k$, in which the independent variable k takes on only a finite number of values, the integers from 1 to n. In the same way, each of the sets

$(a_{i1}, a_{i2}, \cdots, a_{im})$ may be regarded as a function $a_i(k)$. The x's then are the coefficients in a linear combination of these functions,

$$w_k = w(k) = x_1 a_1(k) + x_2 a_2(k) + \cdots + x_m a_m(k),$$

and are to be determined so that

$$\sum_{k=1}^{n} (b_k - w_k)^2$$

shall be as small as possible. The sum of squares this time measures the discrepancy between the approximation and the function approximated.

The form of the sum suggests a geometric interpretation, in which the functions $b(k)$, $w(k)$ are represented by points in n-dimensional space, with coördinates $(b_1, b_2, \cdots, b_n)$ and $(w_1, w_2, \cdots, w_n)$ respectively, and $\sum (b_k - w_k)^2$ is the square of the distance between these points. By an extension of the same idea, the functions $f(x)$, $\varphi(x)$ of the first paragraph are thought of as corresponding to points in a space of infinitely many dimensions, with $\int [f(x) - \varphi(x)]^2\, dx$ as the square of the distance between them. This definition of distance is the beginning of a systematic geometry of function space.

Another hint of geometric analogy which is present from the beginning consists in the recurring use of the term *orthogonal*, two functions being called orthogonal to each other over an interval when the integral of their product over the interval is zero. While it is worthy of some emphasis that orthogonality of the functions which form the basis of the approximation is a convenience rather than a necessity, the notion of orthogonality is inseparably associated with the least-square condition in another way.

Let the functions $f(x)$, $p_1(x)$, $\cdots$, $p_m(x)$ be supposed continuous, for simplicity of illustration. (It would be sufficient that they be integrable together with their squares, provided it is assumed that the p's are *properly* independent, in the sense that every linear combination of them containing a non-

vanishing coefficient is different from zero *over a set of positive measure*; this condition is of course satisfied automatically if they are linearly independent and continuous.) *In order that $c_1, \cdots, c_m$ be the coefficients giving the least-square approximation, it is necessary and sufficient that $f(x) - \varphi(x)$ be orthogonal to each of the functions $p_i(x)$,* or in other words *it is necessary and sufficient that $f(x) - \varphi(x)$ be orthogonal to every linear combination of $p_1(x), \cdots, p_m(x)$.* This can be verified algebraically, without a detailed examination of the necessary and sufficient conditions for a minimum in the calculus.

Let $\psi(x)$ be an arbitrary linear combination of the p's with at least one non-vanishing coefficient. Then

$$\omega(x) = \varphi(x) + h\,\psi(x)$$

is a linear combination of the p's, for any value of h, and every linear combination different from φ can be written in this form by a suitable choice of ψ. Let

$$I = \int_a^b [f(x) - \varphi(x)]^2\,dx,$$

$$J = \int_a^b [f(x) - \omega(x)]^2\,dx = \int_a^b \{[f(x) - \varphi(x)] - h\,\psi(x)\}^2\,dx,$$

$$R = \int_a^b [f(x) - \varphi(x)]\,\psi(x)\,dx, \qquad S = \int_a^b [\psi(x)]^2\,dx > 0.$$

Then $J = I - h(2R - hS)$. If $R \neq 0$, h can be chosen so as to make $J < I$ (for example by taking $h = R/S$), contrary to the supposition that φ is the minimizing function, while if $R = 0$, $J = I + h^2 S > I$ for every $h \neq 0$.

The conditions $\int [f(x) - \varphi(x)]\,p_i(x)\,dx = 0$, $i = 1, 2, \cdots, m$, are equivalent to the m linear equations

$$\sum_{j=1}^{m} c_j \int_a^b p_j(x)\,p_i(x)\,dx = \int_a^b f(x)\,p_i(x)\,dx$$

for determining the m c's. It has been shown that these equations must be satisfied if the least-square problem is to

be solved, and that the problem will be solved if the equations are satisfied. For the moment there is still a question whether the equations have a solution. The answer is immediate, however; the left-hand members are independent of the function $f(x)$, the least-square problem obviously has the unique solution $c_1 = c_2 = \cdots = c_m = 0$ in the particular case $f(x) = 0$, and consequently the determinant of the coefficients must be different from zero. If the p's were not linearly independent, the problem of approximation could still be solved in terms of a linearly independent subset of them, and this would be at the same time a solution in terms of the original set of p's, but the solution in terms of the latter set as a whole would not be unique.

Similar reasoning is applicable to the problem of the second paragraph of the chapter. Integrals are to be replaced by sums throughout, and in particular the property of orthogonality of two sets of numbers is expressed by the vanishing of the sum of the products of corresponding numbers of the two sets. The hypothesis of linear independence of the p's corresponds to the condition that the matrix of the coefficients a_{ik} be of rank m. The condition of orthogonality characterizing the least-square solution takes the form that *the set of residuals $b_k - w_k$ is orthogonal to each of the m sets of numbers a_{ik}:*

$$\sum_{k=1}^{n} (b_k - w_k)\, a_{ik} = 0.$$

The ith equation of this set may be constructed by multiplying each of the given equations by the corresponding coefficient of x_i, and adding the equations thus obtained. Written out at length, the new equations have the form

$$\left(\sum_{k=1}^{n} a_{ik}\, a_{1k}\right) x_1 + \left(\sum_{k=1}^{n} a_{ik}\, a_{2k}\right) x_2 + \cdots + \left(\sum_{k=1}^{n} a_{ik}\, a_{mk}\right) x_m$$
$$= \sum_{k=1}^{n} a_{ik}\, b_k, \qquad i = 1, 2, \cdots, m.$$

They are the well-known "normal equations" for the solution of the problem of least-square adjustment.

In certain cases the necessary condition of orthogonality can readily be translated into another familiar form. Consider once more the case of functions of a continuous variable x. Let $f(x)$ and $\varrho(x)$ be given functions for $a \leq x \leq b$, for simplicity continuous, and let $\varrho(x)$ be non-negative and not identically zero over the interval, and let $P_n(x)$ be the polynomial of the nth degree which minimizes the integral

$$\int_a^b \varrho(x)\,[f(x) - P_n(x)]^2\, dx.$$

Then the remainder $f(x) - P_n(x)$, if not identically zero wherever $\varrho(x) \neq 0$, must change sign at least $n+1$ times in the interval (a, b). If $[\varrho(x)]^{1/2}$ is denoted by $q(x)$, the integral to be minimized is the same as $\int [q(x) f(x) - q(x) P_n(x)]^2\, dx$, and the problem is that of approximating $q(x) f(x)$ by a linear combination of the functions $q(x),\ x q(x), \cdots, x^n q(x)$. By the general proposition obtained above, it is necessary that $q(x) f(x) - q(x) P_n(x)$ be orthogonal to every linear combination of the functions $x^i q(x)$, or, in other words, orthogonal to $q(x)\, Q_n(x)$, if $Q_n(x)$ is an arbitrary polynomial of the nth degree; in symbols, since $[q(x)]^2 = \varrho(x)$,

$$\int_a^b \varrho(x)\,[f(x) - P_n(x)]\, Q_n(x)\, dx = 0.$$

If $f(x) - P_n(x)$ had not more than n changes of sign, it would be possible to construct a $Q_n(x)$ having the same sign as $f(x) - P_n(x)$ at all points where $f(x) - P_n(x) \neq 0$, and this would make the integral positive, unless $\varrho(x) = 0$ wherever $f(x) - P_n(x) \neq 0$. The condition that $P_n(x)$ be such as to give the remainder the requisite number of changes of sign is of course not sufficient for the solution of the least-square problem, since it is satisfied by any polynomial differing sufficiently little from the minimizing polynomial. A corresponding formulation is possible in the case of approximation by trigonometric sums, with or without a weight function.

The least-square problem, in general as well as in the special cases discussed in the preceding paragraph, is that

of choosing from a given linear family of functions the particular one which is closest to another given function, according to the measure of discrepancy specified. The geometric interpretation that has been suggested gains in clearness if each function, instead of being represented merely by a point in a space of an appropriate number of dimensions, is also represented alternatively by the vector from the origin to the point in question. The linear family of functions then corresponds to a linear spread in the geometric picture, and the function of closest approximation corresponds to the point of this spread whose distance from the point representing the function to be approximated is as small as possible; the residual function $f(x) - \varphi(x)$ (or $b(k) - u(k)$) can be regarded as the vector from $\varphi(x)$ to $f(x)$ (or from $u(k)$ to $b(k)$); and the property of "orthogonality" characterizing it is associated with the fact that the shortest distance from a plane spread to a point outside it is perpendicular to the spread. The geometric terminology thus acquires additional plausibility.

2. The general notion of angle; geometric interpretation of coefficients of correlation

An obvious further step is to proceed from the notion of orthogonality to a general notion of angle in function space. As will be seen presently, the definition of angle is already implicit in that of distance, if the frame of Euclidean geometry is to be fitted consistently to functional relations, though the question as to the possibility of carrying the idea through systematically and without danger of internal contradiction still calls for some elucidation.

It will be well first to adopt a common notation for dealing simultaneously with functions of a continuous variable and functions of a discrete subscript. If $x(t)$, $y(t)$ are continuous functions of t for $a \leq t < b$, let

$$(x \cdot y) = \int_a^b x(t)\, y(t)\, dt.$$

If $x_1, x_2, \cdots, x_n$; $y_1, y_2, \cdots, y_n$ are two functions of an index k which ranges from 1 to n, let

$$(x \cdot y) = \sum_{k=1}^{n} x_k y_k.$$

The quantity $(x \cdot y)$ is the inner or scalar product of the two functional vectors. In particular, $(x \cdot x) = \sum x_k^2$ or $\int x^2 \, dt$, etc. The condition for orthogonality of x and y is that $(x \cdot y) = 0$. A third case worthy of explicit mention is that of infinite sequences $x_1, x_2, \cdots; y_1, y_2, \cdots,$ such that $\sum x_k^2$ and $\sum y_k^2$ are convergent. The interpretation then is that

$$(x \cdot y) = \sum_{k=1}^{\infty} x_k y_k.$$

All three are of course isolated particular cases from the point of view of general analysis.

Let $x(t)$, $y(t)$, or $x(k)$, $y(k)$, be two functions represented by the points P, Q respectively, and let O be the origin, corresponding to a function which vanishes identically. By the convention already adopted, the sides of the triangle OPQ are

$$OP^2 = (x \cdot x), \qquad OQ^2 = (y \cdot y),$$
$$PQ^2 = ((y - x) \cdot (y - x)) = (y \cdot y) - 2(x \cdot y) + (x \cdot x).$$

If the angle POQ is denoted by θ, application of the law of cosines to the triangle gives

$$PQ^2 = OP^2 + OQ^2 - 2\,OP \cdot OQ \cos\theta,$$

whence $2\,OP \cdot OQ \cos\theta = 2(x \cdot y)$, and

$$\cos\theta = \frac{(x \cdot y)}{(x \cdot x)^{1/2} (y \cdot y)^{1/2}}.$$

When the independent variable is $k = 1, 2, \cdots, n$, the formula becomes

$$\cos\theta = \frac{\sum x_k y_k}{(\sum x_k^2)^{1/2} (\sum y_k^2)^{1/2}}.$$

This is recognized immediately as a fundamental formula in the theory of statistics. *It is the coefficient of correlation*

between the variables x and y, if these have been reduced to the form of deviations from their respective arithmetic means, or, in other words, if $\sum x_k = \sum y_k = 0$. Subsequent pages will show that the geometry of function space throws much light on the structure of more complicated formulas of correlation.

It becomes important to inquire more closely as to the logical basis for the identification of analysis and geometry. The essentials of such a basis are implicit in the considerations leading up to the proof of Theorem III in Chapter III. To repeat in the present notation what is needed for the purpose in hand, restricting attention at first to a two-dimensional spread, let x, y be any two functions which are linearly independent over one of the ranges specified in the third paragraph preceding (or over some other appropriate range); if the range is an interval, let it be supposed for convenience that the functions are continuous. Let

$$\xi_1 = x, \qquad \eta_1 = y - \xi_1 \frac{(\xi_1 \cdot y)}{(\xi_1 \cdot \xi_1)}.$$

It follows immediately from this definition that $(\xi_1 \cdot \eta_1) = 0$. By the hypothesis of linear independence, furthermore, it is certain that η_1 is not identically zero. If

$$\xi = \xi_1 / (\xi_1 \cdot \xi_1)^{1/2}, \qquad \eta = \eta_1 / (\eta_1 \cdot \eta_1)^{1/2},$$

the functions ξ, η are orthogonal to each other: $(\xi \cdot \eta) = 0$; and they also satisfy the condition that $(\xi \cdot \xi) = (\eta \cdot \eta) = 1$. They are linear combinations of x and y, and it is seen at once that the determinant of the coefficients is different from zero, so that x and y conversely can be expressed as linear combinations of ξ and η.

The functions x and y being given, let ξ and η now be any pair of functions such that x and y are linearly expressible in terms of ξ and η, and such that

$$(\xi \cdot \eta) = 0, \qquad (\xi \cdot \xi) = (\eta \cdot \eta) = 1.$$

It has been shown that such functions ξ, η can be constructed. There are infinitely many pairs satisfying the requirements; the original ξ, η can be replaced by $\alpha_1\xi+\beta_1\eta$, $\alpha_2\xi+\beta_2\eta$, if α_1, β_1, α_2, β_2 are any constants such that

$$\alpha_1\alpha_2+\beta_1\beta_2=0, \qquad \alpha_1^2+\beta_1^2=\alpha_2^2+\beta_2^2=1.$$

Furthermore, the requirements can be satisfied even if x and y are not linearly independent; it suffices then to express x, y in terms of two linearly independent functions x_1, y_1, and to construct functions ξ, η as above with x_1 and y_1 in place of x and y.

In terms of a chosen pair of normalized orthogonal functions ξ, η, let

$$x=a_1\xi+b_1\eta, \qquad y=a_2\xi+b_2\eta.$$

Then

$$(x\cdot x)=a_1^2+b_1^2, \qquad (y\cdot y)=a_2^2+b_2^2,$$
$$((y-x)\cdot(y-x))=(a_2-a_1)^2+(b_2-b_1)^2.$$

If the functions x, y are thought of as corresponding to the points P, Q, with the coördinates (a_1, b_1) and (a_2, b_2) respectively in a rectangular coördinate system, while O is the origin, the quantities $(x\cdot x)$, $(y\cdot y)$, and $((y-x)\cdot(y-x))$ are the squares of the distances OP, OQ, and PQ. The cosine of the angle POQ is

$$\frac{a_1a_2+b_1b_2}{(a_1^2+b_1^2)^{1/2}(a_2^2+b_2^2)^{1/2}}=\frac{(x\cdot y)}{(x\cdot x)^{1/2}(y\cdot y)^{1/2}}.$$

More generally, if

$$u=\lambda_1x+\mu_1y=A_1\xi+B_1\eta, \qquad v=\lambda_2x+\mu_2y=A_2\xi+B_2\eta$$

are any linear combinations of x and y, and if R, S are the corresponding points (A_1, B_1), (A_2, B_2), then $(u\cdot u)$, $(v\cdot v)$, and $((v-u)\cdot(v-u))$ are the squares of the distances OR, OS, and RS, and the quantity $(u\cdot v)$ is the scalar product of the vectors OR, OS. Every linear combination $\lambda x+\mu y$ corresponds to a definite point with coördinates $(\lambda a_1+\mu a_2, \lambda b_1+\mu b_2)$; if x and y are linearly independent, a one-to-one correspondence

is established between functions linearly dependent on them and the points of a plane, in such a way that there is an actual identity between the measures of distance and angle and the quantities which were associated with them originally by analogy. So far, moreover, there is no need of any space of more than two dimensions.

The correspondence being once established, the analytical relations expressing geometric facts are implied with logical conclusiveness by the geometric facts themselves. Consider for example the problem of determining a constant λ to minimize $((y-\lambda x)\cdot(y-\lambda x))$, when x and y are given. If x and y are interpreted as statistical variables, representing deviations from the respective arithmetic means, λ is the coefficient of regression of y on x. The points corresponding to x and y being P and Q, as before, the function λx is represented by a point M on the line OP, and $y-\lambda x$ corresponds to the vector MQ, in the definite sense that when $y-\lambda x$ is expressed as a linear combination of ξ and η, the coefficients of ξ and η are the components of the vector. It is clear from the geometric figure not only that MQ must be perpendicular to OP, or in other words that $y-\lambda x$ must be orthogonal to x, but also that the value of λ which accomplishes the purpose is given by

$$\lambda\cdot OP = OM = OQ\cos\theta,$$

$$\lambda = \frac{OQ}{OP}\cos\theta = \frac{(y\cdot y)^{1/2}}{(x\cdot x)^{1/2}}\cdot\frac{(x\cdot y)}{(x\cdot x)^{1/2}(y\cdot y)^{1/2}} = \frac{(x\cdot y)}{(x\cdot x)},$$

where θ is the angle POQ, and furthermore that the minimum value of $((y-\lambda x)\cdot(y-\lambda x))$ is

$$MQ^2 = (OQ\sin\theta)^2 = (y\cdot y)(1-\cos^2\theta) = (y\cdot y)(1-r^2),$$

if $\cos\theta$, interpreted as a coefficient of correlation, is denoted by r.

For the geometry of three (linearly independent) functions x, y, z, let ξ_1, η_1, ξ, η be defined as before, and let

$$\zeta_1 = z - \xi_1\frac{(\xi_1\cdot z)}{(\xi_1\cdot\xi_1)} - \eta_1\frac{(\eta_1\cdot z)}{(\eta_1\cdot\eta_1)}, \qquad \zeta = \zeta_1/(\zeta_1\cdot\zeta_1)^{1/2}.$$

Then ξ, η, ζ satisfy the conditions

$$(\xi \cdot \eta) = (\xi \cdot \zeta) = (\eta \cdot \zeta) = 0, \quad (\xi \cdot \xi) = (\eta \cdot \eta) = (\zeta \cdot \zeta) = 1,$$

and x, y, z are linearly expressible in terms of them. There are infinitely many sets of functions satisfying these same conditions, and serving equally well for the representation of x, y, z, any such set being expressible in terms of the particular set ξ, η, ζ by means of the coefficients of an orthogonal linear transformation in three variables. Any linear combination $u = \lambda x + \mu y + \nu z$ can be expressed in the form $A\xi + B\eta + C\zeta$, and can thus be put in correspondence with a point (A, B, C), the square of whose distance from the origin is the quantity $(u \cdot u) = A^2 + B^2 + C^2$. If u and v are two such linear combinations, corresponding to the points P and Q, the cosine of the angle POQ is $(u \cdot v)/[(u \cdot u)(v \cdot v)]^{1/2}$.

The representation of functions by points or vectors is particularly convenient for the visualization of coefficients of partial and double correlation. Let x, y, z be three given functions, corresponding to points P, Q, R in three-dimensional space. If x, y, z are sets of deviations from arithmetic means, so that the statistical terminology is appropriate, the coefficient of partial correlation between x and y, when z is held fast, is the coefficient of correlation between $x - \lambda z$ and $y - \mu z$, where λ and μ are the regression coefficients of x on z and of y on z respectively. The function $x - \lambda z$ is a linear combination of x and z, orthogonal to z; its geometric counterpart is a vector in the plane POR, perpendicular to OR. Similarly, $y - \mu z$ is represented by a vector perpendicular to OR, and lying in the plane QOR. The angle between these vectors measures the dihedral angle formed by the two planes. So the coefficient of partial correlation in question is the cosine of the dihedral angle. Let P_1, Q_1, R_1 be the points in which the rays OP, OQ, OR pierce the unit sphere about the origin as center. In the spherical triangle $P_1 Q_1 R_1$, let α, β, γ be the measures of the angles P_1, Q_1, R_1 respectively, and let a, b, c be the sides opposite these angles. Let r_{12}, r_{13}, r_{23} respectively be the ordinary cor-

relation coefficients of x and y, x and z, and y and z; let $r_{12.3}$ be the partial correlation coefficient which has just been discussed, and let $r_{13.2}$ and $r_{23.1}$ be the other coefficients of partial correlation. It has been seen that $r_{12.3} = \cos\gamma$. In the same way, $r_{13.2} = \cos\beta$, $r_{23.1} = \cos\alpha$, while $r_{12} = \cos c$, $r_{13} = \cos b$, $r_{23} = \cos a$. By the law of cosines.

$$\cos\gamma = \frac{\cos c - \cos a \cos b}{\sin a \sin b},$$

which means that

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{[(1-r_{13}^2)(1-r_{23}^2)]^{1/2}}.$$

This is the standard formula expressing a coefficient of partial correlation in terms of ordinary correlation coefficients. The coefficients $r_{13.2}$, $r_{23.1}$ of course have corresponding expressions. The inverse formulas

$$r_{12} = \frac{r_{12.3} + r_{13.2}\, r_{23.1}}{[(1-r_{13.2}^2)(1-r_{23.1}^2)]^{1/2}},$$

etc., are similarly obtained from the polar triangle.

For the definition of one of the coefficients of double correlation, parameters λ, μ are to be determined so as to minimize the quantity $((x - \lambda y - \mu z) \cdot (x - \lambda y - \mu z))$, or in other words to give the least-square approximation of x by means of a linear combination of y and z. This requires that $x - \lambda y - \mu z$ be orthogonal to both y and z; geometrically, the point M representing the combination $\lambda y + \mu z$ is the foot of the perpendicular from P on the plane QOR. Then $r_{1.23}$. the coefficient of double correlation between x and the pair of variables y and z, is the coefficient of simple correlation between x and $\lambda y + \mu z$, the cosine of the angle MOP between OP and the plane QOR. Let this angle be denoted by h. It is measured by the arc $P_1 S_1$, if S_1 is the foot of the altitude from P_1 to the side $Q_1 R_1$ in the spherical triangle. Hence h may be calculated as a side of the right triangle $R_1 S_1 P_1$, by the formula

$$\sin h = \sin b \sin\gamma.$$

By substitution from the relations

$$\sin^2 b = 1 - r_{13}^2,$$

$$\sin^2 \gamma = 1 - r_{12.3}^2 = \frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2 r_{12} r_{13} r_{23}}{(1 - r_{13}^2)(1 - r_{23}^2)}$$

it is found that

$$r_{1.23} = \cos h = (1 - \sin^2 h)^{1/2} = \frac{(r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23})^{1/2}}{(1 - r_{23}^2)^{1/2}}.$$

There are corresponding formulas for the other double correlation cofficients $r_{2.13}$ and $r_{3.12}$.

The same figure may be used to obtain formulas for the coefficients λ and μ, the partial regression coefficients of x on y and z. Let lines be drawn through M parallel to OR and OQ, meeting OQ and OR at K and L respectively, to form a parallelogram $OKML$. The vectors OK and OL, constituting a resolution of OM into components collinear with OQ and OR, represent separately the terms λy and μz. On the surface of the sphere, let the arcs $R_1 S_1$ and $S_1 Q_1$ be denoted by d and e, so that d measures the angle LOM, while e measures the angle MOK and its equal OML. (The formulation is adapted throughout to the case in which the point M falls within the angle QOR, so that S_1 is interior to the arc $Q_1 R_1$.) In the plane triangle LOM, having two of its angles equal to d and e respectively, the third angle, at L, is the supplement of $d + e$. But the arcs d and e on the sphere make up the side a of the original spherical triangle. So the law of sines in the plane triangle gives

$$\frac{OL}{OM} = \frac{\sin e}{\sin (d + e)} = \frac{\sin e}{\sin a}.$$

In the right spherical triangle $P_1 S_1 Q_1$, on the other hand,

$$\sin e = \frac{\sin c \cos \beta}{\cos h},$$

so that

$$OL = \frac{OM \sin c \cos \beta}{\sin a \cos h},$$

while $OL = \mu \cdot OR$, and $OM = OP \cos h$. Hence

$$\mu = \frac{OL}{OR} = \frac{OP \sin c}{OR \sin a} \cos \beta,$$

or with the use of a conventional notation, presently to be explained at greater length, for regression coefficients and standard errors of estimate, as well as that already used for coefficients of correlation,

$$b_{13.2} = \frac{\sigma_{1.2}}{\sigma_{3.2}} r_{13.2}.$$

Similarly,

$$b_{12.3} = \lambda = \frac{OP \sin b}{OQ \sin a} \cos \gamma = \frac{\sigma_{1.3}}{\sigma_{2.3}} r_{12.3}.$$

The results of the preceding paragraph can be obtained by a possibly less straightforward but more exclusively geometrical method, making no use of the spherical triangle or of spherical trigonometry as such. Let the letters O, P, Q, R, M, K, L have the same signification as before. The letters a, b, c, α, β, γ retain their previous meanings as measures of the face angles and dihedral angles of the trihedral angle $O\text{-}PQR$. Let H be the foot of the perpendicular from M on OQ, and J the foot of the perpendicular from R on OQ. The triangles KMH and ORJ in the plane QOR are similar, since KM is parallel to OR (having been so constructed), MH is parallel to RJ (both being perpendicular to OQ), and KH and OJ are collinear. Furthermore, KM and OL are opposite sides of a parallelogram. So

$$\frac{OL}{OR} = \frac{KM}{OR} = \frac{MH}{RJ}.$$

But the length of RJ, the perpendicular from R on OQ, is $OR \sin a$. Also, PH is perpendicular to OQ, since MP, being perpendicular to the plane QOR, is perpendicular to the line OQ, and OQ, being perpendicular to MP and MH, is perpendicular to their plane and to the line PH in that plane; hence MHP is the measure of the dihedral angle β, so that $MH = HP \cos \beta$ (as PMH is a right angle), while HP

$= OP \sin c$, making $MH = OP \sin c \cos \beta$. The resulting expression for μ is the same as before.

The preceding calculations are not restricted in substance to the case of statistical variables, but can be formulated without the terminology of correlation, and are then applicable to any functions x, y, z coming under the original hypotheses. It is to be emphasized also that for the time being no use has been made of space of more than three dimensions; the geometry is the actual geometry of experience.

3. Coefficients of correlation in an arbitrary number of variables

For dealing with relations of higher complexity it will be convenient to modify the notation somewhat. When there are m functions to be considered, let them be denoted by $x_1, x_2, \cdots, x_m$. The case of primary interest will be that of statistical variables, measured from an arithmetic mean in each case; i. e., x_i will stand for a set of numbers $x_{i1}, x_{i2}, \cdots, x_{in}$, subject to the condition that $x_{i1} + x_{i2} + \cdots + x_{in} = 0$. Apart from technical notation and terminology, however, the work will be valid for functions of any of the types previously considered. As an additional item of notation, $(x \cdot x)$ will be abbreviated to $((x))^2$.

The reduction of general coefficients of correlation and coefficients of regression to expressions in terms of coefficients of lower order depends on the following fundamental proposition:

Let $\lambda_2, \cdots, \lambda_m$ be determined so as to minimize

$$((x_1 - \lambda_2 x_2 - \cdots - \lambda_m x_m))^2,$$

and with these values of the λ's, let

$$\omega = x_1 - \lambda_2 x_2 - \cdots - \lambda_m x_m.$$

Let $\mu_3, \cdots, \mu_m$ and $\nu_3, \cdots, \nu_m$ be determined so as to minimize $((x_1 - \mu_3 x_3 - \cdots - \mu_m x_m))^2$ and $((x_2 - \nu_3 x_3 - \cdots - \nu_m x_m))^2$, and let

$$\varphi = x_1 - \mu_3 x_3 - \cdots - \mu_m x_m, \quad \psi = x_2 - \nu_3 x_3 - \cdots - \nu_m x_m.$$

Let A be determined so as to minimize $((\varphi - A\psi))^2$. *Then*

$$\omega = \varphi - A\psi.$$

The proof depends on the still more fundamental fact, immediately deducible from the definition of orthogonality, that *if one function is orthogonal to each of two other functions, it is orthogonal to every linear combination of them.*

It is at once apparent from the definitions of φ and ψ that $\varphi - A\psi$ is a linear combination of the x's, of the same form as ω, the coefficient of x_1 being unity in each case; the question at issue is that of the identity of the remaining coefficients.

By a theorem discussed early in the chapter, a (necessary and) sufficient condition characterizing the coefficients in ω is that ω be orthogonal to each of the functions $x_2, \cdots, x_m$.

By the same theorem, φ is orthogonal to each of the functions $x_3, \cdots, x_m$, and ψ is likewise orthogonal to each of these functions. Since x_3 is orthogonal to φ and to ψ, it is orthogonal to the combination $\varphi - A\psi$. The same is true of $x_4, \cdots, x_m$. In other words, *$\varphi - A\psi$ is orthogonal to each of the functions $x_3, \cdots, x_m$.*

By one more application of the theorem, $\varphi - A\psi$ is orthogonal to ψ. But x_2 can be expressed in the form $x_2 = \psi + v_3 x_3 + \cdots + v_m x_m$. Consequently, being orthogonal to $\psi, x_3, \cdots, x_m$, *$\varphi - A\psi$ is orthogonal also to x_2.*

The identity of $\varphi - A\psi$ with ω is thus established.

For $m = 3$, the proposition is equivalent to a relation of perpendiculars which is important in deriving the formulas of spherical trigonometry: if OP, OQ, OR are three rays issuing from O, if N is the foot of the perpendicular from P on the line OR, if a line is drawn through N in the plane QOR perpendicular to OR, and if M is the foot of the perpendicular from P on this line, then MP is perpendicular to the plane QOR. This figure yields the equation $\sin h = \sin b \sin \gamma$, which was used in obtaining the formula for a coefficient of double correlation. Essentially the same configuration appeared also in the discussion of partial regression coefficients.

For the statistical application with an arbitrary value of m a somewhat elaborate notation is appropriate. Let the function ω, as originally defined by the formula $x_1 - \lambda_2 x_2 - \cdots - \lambda_m x_m$ according to the least-square criterion, be denoted by $x_{1.23\cdots m}$. It may be spoken of as the *residual of* x_1 *with respect to* $x_2, x_3, \cdots, x_m$. (The order of the subscripts $2, 3, \cdots, m$ among themselves is clearly immaterial.) Similarly, φ and ψ, the residuals of x_1 and of x_2 with respect to $x_3, \cdots, x_m$, are to be denoted by $x_{1.3\cdots m}$ and $x_{2.3\cdots m}$. The *standard deviation* σ_k of any one of the original variables x_k is defined by the equation $\sigma_k^2 = ((x_k))^2/n$. (The presence of the denominator n is a mere matter of definition, as far as the present discussion is concerned, since the equations will involve only *ratios* of standard deviations; except for the statistical interpretation, the (positive) square root of $((x))^2$ itself may be used in place of σ_k.) The correspondingly defined standard deviation of such a residual as $x_{1.23\cdots m}$, the *standard error of estimate of* x_1 *in terms of* $x_2, \cdots, x_m$, is denoted by $\sigma_{1.23\cdots m}$. The *partial regression coefficients* $\lambda_2, \lambda_3, \cdots, \lambda_m$ are represented by $b_{12.34\cdots m}, b_{13.24\cdots m}, \cdots, b_{1m.23\cdots m-1}$. The first subscript of each of the b's is that of the variable approximated, the second is that of the particular variable to which the coefficient in question is attached in the regression formula, and the other subscripts, the order of which among themselves is without significance, are those of the remaining variables. In the same way, $\mu_3, \cdots, \mu_m$ are to be replaced by $b_{13.4\cdots m}, \cdots, b_{1m.3\cdots m-1}$, and $\nu_3, \cdots, \nu_m$ by $b_{23.4\cdots m}, \cdots, b_{2m.3\cdots m-1}$. The coefficient of correlation between φ and ψ, or, in the present notation, between $x_{1.3\cdots m}$ and $x_{2.3\cdots m}$, is the *coefficient of partial correlation between* x_1 *and* x_2 *when* $x_3, \cdots, x_m$ *are held fast*, and is represented by $r_{12.3\cdots m}$.

In the new notation, the general proposition about the identity of the functions previously called ω and $\varphi - A\psi$ asserts that $x_{1.23\cdots m}$ is the same as the residual of $x_{1.3\cdots m}$ with respect to $x_{2.3\cdots m}$. It is important to bear in mind however that its essential content is independent of the

number of variables and their individual designations, and this essential content is more adequately though less concisely expressed by saying:

The residual of a given function with respect to a set of functions may be obtained by calculating successively the residuals of the given function and of any chosen function of the set with respect to the remaining functions of the set, and then taking the residual of the former of these residuals with respect to the latter.

An ealier paragraph contained a derivation of the value of the simple regression coefficient minimizing the expression $((y - \lambda x)\cdot(y - \lambda x))$. The formula of that paragraph which reads

$$\lambda = \frac{(y\cdot y)^{1/2}}{(x\cdot x)^{1/2}}\cdot\frac{(x\cdot y)}{(x\cdot x)^{1/2}\,(y\cdot y)^{1/2}}$$

becomes in the present notation $b_{21} = (\sigma_2/\sigma_1)\,r_{12}$. Interchange of the variables gives $b_{12} = (\sigma_1/\sigma_2)\,r_{12}$; the simple coefficient of correlation is symmetrical in its two variables, so that $r_{21} = r_{12}$, while b_{21} and b_{12} are different. Incidentally it appears that $r_{12} = (b_{12}\,b_{21})^{1/2}$.

Corresponding formulas for partial regression coefficients can be obtained immediately. It is apparent on inspection that while the coefficients of $x_3, \cdots, x_m$ in the function $x_{1.23\cdots m} = \omega$, as expressed in the form $\varphi - A\psi$, are combinations of the μ's and the ν's, *the coefficient* $\lambda_2 = b_{12.3\cdots m}$ *is merely* A. But this A is the simple coefficient of regression of φ with respect to ψ, and as such is expressible in terms of the standard deviations of φ and ψ and the correlation between them. The standard deviations of φ and ψ are $\sigma_{1.3\cdots m}$ and $\sigma_{2.3\cdots m}$ respectively, and their coefficient of correlation is $r_{12.3\cdots m}$. Consequently

$$b_{12.3\cdots m} = \frac{\sigma_{1.3\cdots m}}{\sigma_{2.3\cdots m}}\,r_{12.3\cdots m}.$$

By interchange of subscripts—in other words, by consideration of the coefficient of regression of $x_{2.3\cdots m}$ with respect to $x_{1.3\cdots m}$—it is found that

$$b_{21.3\cdots m} = \frac{\sigma_{2.3\cdots m}}{\sigma_{1.3\cdots m}} r_{12.3\cdots m};$$

the definition of the correlation coefficient is symmetrical with respect to the first two subscripts. Combination of the equations for the b's gives

$$r_{12\,3\cdots m} = (b_{12\,3\cdots m}\, b_{21.3\cdots m})^{1/2}.$$

In the earlier paragraph to which reference was made above the minimum of $((y-\lambda x)\cdot(y-\lambda x))$ was evaluated in the form $(y\cdot y)(1-r^2)$. In the new notation this result is expressed by the equation $\sigma_{2.1}^2 = \sigma_2^2(1-r_{12}^2)$. The corresponding equation with subscripts reversed is $\sigma_{1.2}^2 = \sigma_1^2(1-r_{12}^2)$. Applied to the standard deviation of $\omega = \varphi - A\psi$, regarded as the residual of φ with respect to ψ, it becomes

$$\sigma_{1\,23\cdots m}^2 = \sigma_{1\,3\cdots m}^2(1-r_{12.3\cdots m}^2).$$

Here again it is important to recognize the essential content of the formula, as distinguished from the notation in which it is expressed. Written down successively for cases of increasing complexity, with a particular choice as to the disposition of subscripts each time, it yields

$$\sigma_{1.2}^2 = \sigma_1^2(1-r_{12}^2),$$

$$\sigma_{1.23}^2 = \sigma_{1\,32}^2 = \sigma_{1\,2}^2(1-r_{13.2}^2),$$

$$\sigma_{1.234}^2 = \sigma_{1.423}^2 = \sigma_{1.23}^2(1-r_{14.23}^2),$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\sigma_{1\,23\cdots m}^2 = \sigma_{1\,m23\cdots m-1}^2 = \sigma_{1\,23\cdots m-1}^2(1-r_{1m.23\cdots m-1}^2).$$

and by combination of all these equations

$$\sigma_{1.23\cdots m}^2 = \sigma_1^2(1-r_{12}^2)(1-r_{13\,2}^2)(1-r_{14\,23}^2)\cdots(1-r_{1m.23\cdots m-1}^2).$$

Reverting to an abbreviated notation for certain residuals, but with a modification of that previously employed, let

$$\omega_1 = x_{1.34\cdots m}, \qquad \omega_2 = x_{2.34\cdots m},$$

$$\varphi_1 = x_{1\,4\cdots m}, \qquad \varphi_2 = x_{2\,4\cdots m}, \qquad \varphi_3 = x_{3.4\cdots m}.$$

By the fundamental theorem about residuals, ω_1 may be regarded as the residual of φ_1 with respect to φ_3, and ω_2 is the residual of φ_2 with respect to φ_3. Hence *the coefficient of correlation between ω_1 and ω_2 is the coefficient of partial correlation between φ_1 and φ_2 when φ_3 is held fast*, and is expressed in terms of the simple correlations between φ_1, φ_2 and φ_3 by the formula previously obtained in the discussion of correlations among three variables. If the correlation between ω_1 and ω_2 is denoted for the moment simply by r', and if the correlations between φ_1 and φ_2, φ_1 and φ_3, and φ_2 and φ_3 are called respectively r'_{12}, r'_{13} and r'_{23}, then

$$r' = \frac{r'_{12} - r'_{13} r'_{23}}{[(1 - r'^2_{13})(1 - r'^2_{23})]^{1/2}}.$$

But from the point of view of the dependence of these quantities on $x_1, x_2, \cdots, x_m$,

$$r' = r_{12.34\cdots m}, \quad r'_{12} = r_{12.4\cdots m}, \quad r'_{13} = r_{13.4\cdots m}, \quad r'_{23} = r_{23.4\cdots m},$$

and consequently

$$r_{12.34\cdots m} = \frac{r_{12.4\cdots m} - r_{13.4\cdots m}\, r_{23.4\cdots m}}{[(1 - r^2_{13.4\cdots m})(1 - r^2_{23.4\cdots m})]^{1/2}}.$$

By the last equation any partial correlation coefficient can be calculated in terms of correlation coefficients of lower order, that is to say, correlation coefficients involving a smaller number of variables, and so ultimately in terms of ordinary correlation coefficients.

Let the partial regression coefficients of x_1 with respect to $x_2, x_3, \cdots, x_m$ be denoted once more by $\lambda_2, \lambda_3, \cdots, \lambda_m$, and let $\Phi = \lambda_2 x_2 + \lambda_3 x_3 + \cdots + \lambda_m x_m$. The coefficient of correlation between x_1 and Φ is the *coefficient of multiple correlation between x_1 and the set of variables $x_2, \cdots, x_m$*. It may be represented by $r_{1.23\cdots m}$. If the regression coefficient of x_1 on Φ is L, the square of the standard deviation of $x_1 - L\Phi$ is $\sigma_1^2(1 - r^2_{1.23\cdots m})$. But any constant multiple of Φ is a linear combination of $x_2, \cdots, x_m$, and by the definition

of Φ no other linear combination of $x_2, \cdots, x_m$ can give so good an approximation to x_1 according to the least-square criterion as Φ itself. *Hence it must be that* $L = 1$, and the standard deviation of $x_1 - L\Phi$ is the standard deviation of $x_1 - \Phi$, already denoted by $\sigma_{1\,23 \cdot\cdot m}$, so that

$$\sigma^2_{1\,23\cdot\cdot m} = \sigma^2_1(1 - r^2_{1.23\cdot\cdot m}).$$

Taken in conjunction with an equation previously obtained for $\sigma^2_{1\,23\cdot\cdot m}$, this shows that

$$1 - r^2_{1\,23\cdot\cdot m} = (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23}) \cdots (1 - r^2_{1m\,23\cdot\cdot m-1}),$$

a relation from which the coefficient of multiple correlation can be calculated.

4. The geometry of frequency functions

Apart from the generality which the preceding account of the application of geometry to analysis possesses by virtue of the fact that at the outset the independent variable may be one taking on a finite number, an enumerable infinity, or a continuous infinity of values, its substance can be given still another setting. In the statistical case, if there are for example just two functions concerned, $x_k = x(k)$ and $y_k = y(k)$, the formulas involve the different values of k symmetrically, and are not affected if the n pairs of numbers (x_k, y_k) are permuted, each pair by itself being kept inviolate. That is to say, the independent variable serves only to define the association of a value of x with a value of y, and has no further significance of its own. For dealing with large numbers of observations, the problem may be idealized by supposing that (within limits, perhaps) any value of x may be associated with any value of y, but that some pairs of values (x, y) occur oftener than others, to an extent indicated by a *frequency function* $\varphi(x, y)$, whose integral over any region of the x, y plane measures the frequency of pairs of observations (x, y) falling within that region. If x and y

are measured as deviations from their respective means, this fact is indicated by the conditions

$$\iint x\,\varphi(x, y)\,dx\,dy = 0, \qquad \iint y\,\varphi(x, y)\,dx\,dy = 0,$$

the integrals being extended over the range of definition of φ. The squares of the standard deviations of x and y are then

$$\frac{\iint x^2\,\varphi(x, y)\,dx\,dy}{\iint \varphi(x, y)\,dx\,dy}, \qquad \frac{\iint y^2\,\varphi(x, y)\,dx\,dy}{\iint \varphi(x, y)\,dx\,dy},$$

and the coefficient of correlation between them is

$$\frac{\iint xy\,\varphi(x, y)\,dx\,dy}{\left[\iint x^2\,\varphi(x, y)\,dx\,dy\right]^{1/2}\left[\iint y^2\,\varphi(x, y)\,dx\,dy\right]^{1/2}}.$$

To illustrate the geometry of frequency functions, let the case of three dimensions be chosen. Let $\varphi(x, y, z)$ be a non-negative continuous function of its three arguments, to be regarded as a frequency function for the occurrence of the set of values (x, y, z) for three measured variables. To obviate the necessity of convergence proofs, let it be supposed that φ is different from zero only over a finite domain of three-dimensional space. It may further be supposed without essential loss of generality that the triple integral of φ over the domain where it does not vanish is equal to unity—in abbreviated symbolism, $\int \varphi = 1$—which means that all frequencies are referred to the total number of cases as unit. The assumption that x, y, z are measured from their arithmetic means is equivalent to the set of equations $\int x\varphi = \int y\varphi = \int z\varphi = 0$; no use will be made of these equations, except to justify the statistical terminology employed. The squares of the standard deviations of x, y, z are $\int x^2\varphi$, $\int y^2\varphi$, and $\int z^2\varphi$, since the quantity $\int \varphi$, which

would otherwise appear as denominator, is 1. The coefficient of correlation between x and y is $\int xy\,\varphi \Big/ \left[\left(\int x^2\varphi\right)\left(\int y^2\varphi\right)\right]^{1/2}$, and the other coefficients of correlation are correspondingly defined.

As a first step toward the setting up of a geometrical representation, let auxiliary notation be introduced as follows:

$$X = x, \qquad \left[\int X^2\varphi\right]^{1/2} = \sigma_1, \qquad \int Xy\varphi = A,$$

$$Y = y - (A/\sigma_1^2)X, \quad \left[\int Y^2\varphi\right]^{1/2} = \tau_1, \quad \int Xz\varphi = B, \quad \int Yz\varphi = C,$$

$$Z = z - (B/\sigma_1^2)X - (C/\tau_1^2)Y, \qquad \left[\int Z^2\varphi\right]^{1/2} = \omega_1.$$

In these formulas X, Y, and Z, as well as φ, are to be thought of as functions of x, y, and z, and the integration as performed with regard to these variables. As the functional determinant of X, Y, Z with respect to x, y, z is 1, however, the integrals may equally well be taken with regard to X, Y, Z, if the limits of integration are suitably adjusted, or, what comes to the same thing (since each integrand is identically zero except over a finite region), if the integrals are extended over the whole of space. As a function of X, Y, Z, let $\varphi(x, y, z)$ be denoted by $\varphi_1(X, Y, Z)$. Then, with X, Y, Z as variables of integration,

$$\int XY\varphi_1 = \int XZ\varphi_1 = \int YZ\varphi_1 = 0.$$

Let

$$\xi_1 = X/\sigma_1, \qquad \eta_1 = Y/\tau_1, \qquad \zeta_1 = Z/\omega_1,$$

and let the quantity $\sigma_1\tau_1\omega_1\varphi(x, y, z) = \sigma_1\tau_1\omega_1\varphi_1(X, Y, Z)$, as a function of the new variables, be represented by $\Phi_1(\xi_1, \eta_1, \zeta_1)$. In terms of ξ_1, η_1, ζ_1 as variables of integration, the relations

$$\int \xi_1\eta_1\Phi_1 = \int \xi_1\zeta_1\Phi_1 = \int \eta_1\zeta_1\Phi_1 = 0$$

are satisfied, together with the equations

$$\int \xi_1^2 \Phi_1 = \int \eta_1^2 \Phi_1 = \int \zeta_1^2 \Phi_1 = 1,$$

while $\int\int\int \varphi(x, y, z)\, dx\, dy\, dz = 1$ goes over into

$$\int\int\int \Phi_1(\xi_1, \eta_1, \zeta_1)\, d\xi_1\, d\eta_1\, d\zeta_1 = 1.$$

If $a_1 \xi_1 + b_1 \eta_1 + c_1 \zeta_1$ and $a_2 \xi_1 + b_2 \eta_1 + c_2 \zeta_1$ are any two linear combinations of ξ_1, η_1, ζ_1,

$$\int (a_1 \xi_1 + b_1 \eta_1 + c_1 \zeta_1)(a_2 \xi_1 + b_2 \eta_1 + c_2 \zeta_1)\, \Phi_1 = a_1 a_2 + b_1 b_2 + c_1 c_2.$$

Finally, for the sake of generality, let ξ, η, ζ be any variables expressible in terms of ξ_1, η_1, ζ_1 by means of a (normalized) orthogonal transformation. Let $\Phi_1(\xi_1, \eta_1, \zeta_1) = \Phi(\xi, \eta, \zeta)$. As the determinant of the transformation is ± 1, the equations of the preceding paragraph, including the last one as formulated with the various sets of coefficients involved in the transformation, give

$$\int \xi \eta \Phi = \int \xi \zeta \Phi = \int \eta \zeta \Phi = 0,$$

$$\int \Phi = \int \xi^2 \Phi = \int \eta^2 \Phi = \int \zeta^2 \Phi = 1,$$

the variables of integration now being ξ, η, ζ. The basis of the geometrical interpretation is the possibility of finding ξ, η, ζ as linearly independent linear combinations of x, y, z, so that these relations are satisfied.

The equations expressing ξ, η, ζ in terms of x, y, z manifestly are in fact linearly independent, and x, y, z consequently can be linearly expressed in terms of ξ, η, ζ. Furthermore, any linear combination of x, y, z can be similarly expressed. Let

$$U = \alpha_1 x + \beta_1 y + \gamma_1 z = A_1 \xi + B_1 \eta + C_1 \zeta$$

be any such combination. Then

$$\int\int\int U^2 \varphi\, dx\, dy\, dz = \int\int\int (A_1 \xi + B_1 \eta + C_1 \zeta)^2\, \Phi\, d\xi\, d\eta\, d\zeta$$
$$= A_1^2 + B_1^2 + C_1^2.$$

If

$$V = \alpha_2 x + \beta_2 y + \gamma_2 z = A_2 \xi + B_2 \eta + C_2 \zeta$$

is any second combination of the same form,

$$\int\int\int U V \varphi \, dx \, dy \, dz = A_1 A_2 + B_1 B_2 + C_1 C_2 .$$

Let P be the point with coördinates (A_1, B_1, C_1), and Q the point (A_2, B_2, C_2), with reference to a system of rectangular axes in three dimensions, and let O be the origin. Then $\int U^2 \varphi$ and $\int V^2 \varphi$, the variables of integration being x, y, z, are the squares of the distances OP and OQ, and

$$\int U V \varphi \Big/ \left[\left(\int U^2 \varphi\right)\left(\int V^2 \varphi\right)\right]^{1/2},$$

the coefficient of correlation between U und V, is the cosine of the angle POQ. Through the medium of the equations expressing x, y, z in terms of ξ, η, ζ, each of the variables x, y, z, and every linear combination of them, can be associated with a definite point in three-dimensional space, in such a way that standard deviation and coefficient of correlation have the same sort of geometrical meaning as before. From this beginning the geometrical structure of the theory of correlation can be built up along the lines previously followed.

5. Vector analysis in function space

There is further scope for the application of simple geometric ideas in functional analysis, where the complete picture calls for a geometry of infinitely many dimensions. For a single illustration (discussed by Lévy, *Leçons d'analyse fonctionnelle*, Paris, 1922, pp. 127–128), consider the integral

$$\Omega = \Omega(y) = \int_a^b F(x, y, y') \, dx ,$$

where y is a (suitably) arbitrary function of x, $y = f(x)$, and F is a given function of its arguments. Not to enter into details with regard to questions of continuity, let it be assumed that all the functions that appear in the discussion have as many continuous derivatives as are needed to justify

the operations performed. If y is looked upon as a point in a function space of infinitely many dimensions, or as a vector from the origin to the point, $\Omega(y)$ is a scalar point function in that space. Subject to the appropriate conditions of differentiability, let $\eta(x)$ be an arbitrary function vanishing at a and at b, and h an arbitrary constant. The familiar process of differentiation gives

$$\left[\frac{d}{dh}\,\Omega(y+h\eta)\right]_{h=0} = \int_a^b [\eta\, F_y(x,y,y') + \eta' F_{y'}(x,y,y')]\,dx$$
$$= \int_a^b \eta\left[F_y(x,y,y') - \frac{d}{dx}\,F_{y'}(x,y,y')\right]dx,$$

the last expression resulting from an integration by parts performed on the second term in the brackets, with use of the fact that $\eta(a) = \eta(b) = 0$. Let the function $F_y - (d/dx)F_{y'}$ be denoted by $\varphi(x)$. Then the expression for the derivative may be abbreviated to $\int \eta\,\varphi\,dx$, or, in a notation previously employed, $(\eta\cdot\varphi)$. The variation $h\eta$ may be regarded as an infinitesimal vector increment of the vector y in function space, of geometric magnitude $h(\eta\cdot\eta)^{1/2}$. If the increment of Ω is divided by this quantity, instead of h, passage to the limit gives a *directional derivative* in the direction of the vector η. Its value is $(\eta\cdot\varphi)/(\eta\cdot\eta)^{1/2}$. But this can be written in the form

$$(\varphi\cdot\varphi)^{1/2}\,\frac{(\eta\cdot\varphi)}{(\eta\cdot\eta)^{1/2}\,(\varphi\cdot\varphi)^{1/2}},$$

from which it appears that *the directional derivative in the direction η is equal to the quantity $(\varphi\cdot\varphi)^{1/2}$ multiplied by the cosine of the angle between the vectors η and φ, being greatest when η is collinear with φ. The "functional derivative" φ, considered as a vector in function space, thus has the character of a gradient of the scalar point function Ω.*

The functional derivative can also be obtained formally after the analogy of the ordinary representation of a gradient in terms of a rectangular coördinate system. Let it be supposed now that $y = f(x)$ itself vanishes at the ends of

the interval (a, b), and that it is expanded in a series of normalized orthogonal functions $u_k(x)$, each of which likewise vanishes at the ends of the interval, in the form

$$f(x) = a_1 u_1(x) + a_2 u_2(x) + \cdots.$$

Let it be assumed further that this series admits differentiation term by term. The integral Ω is a function of the infinitely many variables $(a_1, a_2, \cdots)$, the coördinates of the point y with respect to a set of rectangular axes in space of an enumerable infinity of dimensions. The derivative of Ω with regard to a_k is

$$\int_a^b \frac{\partial}{\partial a_k} F(x, y, y')\, dx$$
$$= \int_a^b [F_y(x, y, y')\, u_k(x) + F_{y'}(x, y, y')\, u'_k(x)]\, dx$$
$$= \int_a^b \left[F_y(x, y, y') - \frac{d}{dx} F_{y'}(x, y, y')\right] u_k(x)\, dx$$
$$= \int_a^b \varphi(x)\, u_k(x)\, dx.$$

This is the Fourier coefficient of $\varphi(x)$ with respect to $u_k(x)$. *The expression*

$$\sum_{k=1}^{\infty} \frac{\partial \Omega}{\partial a_k} u_k(x)$$

for a vector having the component $\partial\Omega/\partial a_k$ in the direction of the corresponding coördinate axis is the formal expansion of the functional derivative φ according to the orthogonal system (u_k).

An exposition of other elementary developments of the vector analysis of function space, rendered concrete by the use of theorems on the convergence and degree of convergence of certain expansions in series of orthogonal functions, has been given by the author elsewhere (Annals of Mathematics, (2), vol. 27 (1926), pp. 551–567; Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 641–643).

INDEX OF PRINCIPAL THEOREMS

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V